

New ways with Japan's education?

The Nakasone government would reform the Japanese school system. It should begin where the problems may be simpler — with the universities, much in need of a greater measure of autonomy.

THE Nakasone government is nothing if not courageous. Its decision to embark on a reform of the Japanese educational system (see p.678), or even on a study thereof, is a momentous step, but by no means premature. For decades now, one of the great anomalies of the relationship between Japan and the rest of the industrialized world has been that the economy that has been the most consistently successful should also have been educationally the most backward. From this state of affairs, several mistaken conclusions have been drawn. In Japan, for example, it is often argued that the coexistence of the rigours of life at school and of economic growth must imply a causal relationship — being drilled to work hard at school breeds the compliance and diligence that makes for economic growth. And elsewhere, as economically less successful countries seek to emulate Japan's success, the suspicion is aroused that it may be prudent to emulate the Japanese university pattern in which academic departments are mostly forced to look to industrial partners for substantial research support. The truth, however, is that there are independent reasons for believing that Japanese education is qualitatively deficient in ways that may be even more alarming for Japan's industrial partners and competitors: if Japan can do so well with one hand tied behind its back, what might it not accomplish if the educational system were better suited to the talents of young Japanese?

Even in education, the Japanese achievement should not be underestimated, however. Virtual universal literacy is one obvious benefit, the product of school education. So, by the multiplication of private universities alongside those more directly regulated by the government, is the high participation rate in higher education, more than 30 per cent for women as well as men. In both respects, Japan is the envy of the industrialized world. But the Japanese themselves are now forever asking whether the technical community in Japan is less creative than those elsewhere, or some other version of the question whether the educational system that provides such skills and diligence at work may not also serve to snuff out the spark of originality. Although Mr Nakasone's chief objective may be to relieve students still at school (and their parents) of the pressures to which they are now subjected, the clamour for creativity would be another equally legitimate platform from which to launch a programme of reform.

Examinations

How should reform be planned? The complaint that the system of school education is warped by the requirements of university entrance is probably valid, whence the inevitable temptation to think of ways of decoupling the schools and the universities. A worthwhile programme of reform is almost certain to include some elements with that objective. The danger, in the months that lie ahead, is that the proposed commission to recommend reforms will concentrate on the outward characteristics of the educational system, in particular the prevalence of school examinations, and overlook some of the reasons why the system has grown to be what it is.

One obvious structural deficiency is that the public universities are not nearly as free to conduct their own affairs as they should be. Even the form of a university's curriculum requires not merely the approval of the Ministry of Education but also (formally at

least) that of the Diet. (Thus a decision to set up a new department can be a major undertaking.) Then the discretionary funds with which universities are provided by the government are so meagre (as are the salaries which university teachers are paid) that a department cannot hope to mount an adventurous research programme without help from some outside source. The result is that even the universities most sought after by high-school students, Tokyo and Kyoto, for example, are much less able to direct their own affairs than they should be.

The consequences of these arrangements for the pattern of university research in Japan are well understood but not especially encouraging for academic science. Often, the ministry will generously support particular research projects, most conspicuously the University of Tokyo's own successful space programme. More recently, there has also been welcome support for central facilities to which all universities have access, while modest research grants are more accessible. But there remain many university departments, even in the best-known universities, where worthwhile research cannot be mounted on a sufficient scale. And other university departments are driven into the arms of particular industrial companies to such a degree that research is too narrowly aimed at the solution of industrial problems. While the graduate students emerging from these environments are naturally able to slip easily into productive employment with the sponsors of the research on which they have been engaged, they are by the same token less likely to be the creatively subversive recruits that dynamic industry requires.

Autonomy

So one element of reform should be the provision of a greater measure of autonomy, financial as well as constitutional, to the public universities. That would be good for the universities themselves, the academics who work there and the graduates who emerge from them. What would that accomplish for the armies of high-school students struggling for high grades in their examinations? In the short run, perhaps very little. The trouble with the present pattern of education in Japan is that the best-known universities to which the brightest students aspire are also those which naturally attract most generous public support for research. Moreover, the popular legend that the best-known universities are by definition those from which the best-qualified people emerge is reinforced by the tendency for university departments to select their graduate students from among their own able undergraduates. But if an element of critical and impartial review of university research were more conspicuous in determining the pattern of what universities do, the pre-eminence of the universities now at the top of the pecking order would be less easily taken for granted. And if the possibility that the universities now at the top of the pile might not always be there were appreciated by school-leavers and their teachers, the result would be a better perspective of the importance of high-school education. This, in the long run, is the best insurance against a return of unreasonable elitism in the schools.

The result of such a change could also be a greater diversity within the pattern of the public universities. Broadly speaking, among the score of so of the better-known universities in Japan, there is a single academic goal — that represented by the social influence and prestige of Tokyo and Kyoto. The idea that publicly



supported universities might legitimately seek to differ in character from each other comes into its own only at the other end of the spectrum, at institutions that are essentially vocational in character. The paradoxical result is that there seems no way in which the public university system could at present generate within its own ranks the kind of diversity that gives strength even to the publicly supported universities in the United States (or even within the University of California). And while the private universities in Japan are not merely free but even compelled to respond to a diversity of market forces, their freedom to decide what kinds of institutions they seek to be is more narrowly compromised (which is in no sense to suggest that the private universities are second-rate places but merely that they have to struggle harder to keep alive). □

China's reactors

The US Administration should sell nuclear technology to China only at a diplomatic price.

THE ambition of the United States to sell two reactors to mainland China has been well advertised for the past several months. With the nuclear industry in the doldrums, everywhere, the commercial pressures to arrive at some kind of arrangement are necessarily strong. And since China is almost certain to be one of the principal builders of nuclear power stations in the coming decades, and the only one without first-hand experience of building them, the political as well as commercial prizes to be won by helping with the enterprise from the outset may be very large. The administration's hope that it will be possible to override the requirements of the Atomic Energy Act (see p.677) is therefore understandable.

Even so, the administration should be careful. Among nuclear powers, China is out of the ordinary. So far as can be told, China stands out among the five declared manufacturers of nuclear weapons in its reliance so far on enriched uranium manufactured by means of an isotopic separation plant. The four others make nuclear explosives from reactor plutonium as well. So, in principle, supplying reactors to China without full safeguards on the uses that may be made of the spent nuclear fuel will qualitatively augment China's capacity to manufacture nuclear weapons, and provide more flexibility of design as well. This may not amount to nuclear proliferation in the strict sense, and in any case there is no evidence that China wants to build reactors so as to make nuclear explosives. But on the face of things, what is proposed is contrary to the spirit if not the letter of US legislation on the subject.

A more serious problem is that there is an obvious inconsistency in the proposal that reactors should be built in China without full safeguards so soon after the United States was prevented from supplying spare parts for US reactors built several years ago in India. The obvious source of the difficulty is that the Anti-Proliferation Act distinguishes between nuclear and non-nuclear powers, allowing the supply of nuclear materials to the former (on terms negotiated bilaterally) but not the latter. Curiously, the act seems to provide an incentive for non-nuclear powers to make bombs and thus to qualify for an enhanced status. This should be a reminder to the US administration that President Carter's Anti-Proliferation Act, always offensive to non-nuclear signatories on the Non-Proliferation Treaty as a unilateral imposition, is also shot through with anomalies.

The administration must also calculate its negotiating power more strongly than it has done so far. China has encouragingly joined the UN International Atomic Energy Agency, but that by itself counts for nothing in the safeguards context. What the international community needs is that China should adhere to the Non-Proliferation Treaty as a nuclear power (just as France should), thus providing a means by which it could eventually be drawn into negotiations on the control of nuclear weapons. That is the least that the United States should be asking for in return for supplying reactor technology now. It is not a question of what the law says, but what the administration has the wit to ask for. □

East-West bridges

The death of Piotr Kapitza is a sad blow for international relations.

THE death last week in the Soviet Union of Dr Piotr Kapitza is an especially sad occasion because it breaks one of the few persisting personal links between Soviet science and the West. Kapitza was one of the few Soviet scientists with first-hand experience of working in the West. And while it would be too much to look in the history of relations between the scientific community in the Soviet Union and that elsewhere for signs of a constant beneficial influence by Kapitza, there is no doubt that his presence among the senior members of the Soviet Academy of Sciences has been of great value when East-West relations were sunny.

Kapitza was indeed the archetype of a kind of Soviet scholar that no longer exists. He was born in Kronstadt, the naval base on the Neva downstream from St Petersburg (now Leningrad) in 1894, and visited Britain on an official mission only in 1918, after the Russian Revolution, but by 1921 he had persuaded the Soviet authorities to let him become a research student at the Cavendish Laboratory in Cambridge. By all contemporary accounts, Kapitza's influence on that productive laboratory was profound. Perhaps the most valuable of his innovations at Cambridge was the formation of what afterwards became known as the "Kapitza Club", a kind of dining and discussion club which, from the beginning of Kapitza's second academic year in 1922, became one of the principal means by which the young physics community at Cambridge was stimulated and kept informed of the exciting developments then afoot in mainland Europe. Does Cambridge appreciate even now how much its legendary achievements in the two decades that followed stemmed from the Slav flair for argument that Kapitza brought?

Ernest Rutherford, head of the Cavendish, evidently found Kapitza an appealing character, perhaps even a surrogate son. Kapitza's own research at Cambridge eventually centred on the production of intense magnetic fields, and it seems to be established that Rutherford favoured Kapitza with the assistance needed to recruit the financial support for the construction of the Royal Society Mond Laboratory, opened in 1933 by no less a person than the Prime Minister of Britain, Mr Stanley Baldwin. (Although J. D. Cockcroft, later Lord Cockcroft, an electrical engineer by origin, seems to have helped Kapitza substantially with the design of the equipment for the new laboratory, he complained before his death that his own work on the artificial disintegration of nuclei was, at that time, comparatively starved of funds.) Throughout this period, Kapitza had remained a Soviet citizen and had been allowed to travel freely between Cambridge and the Soviet Union. But at the end of 1934 Kapitza was told that henceforth he must stay in the Soviet Union, with the result that the Soviet Government eventually paid the Cavendish Laboratory £30,000 for the equipment accumulated at the Mond Laboratory. Kapitza (after a period of resistance) became director of the Institute for Physical Problems in Moscow.

There must have been times when the Soviet authorities wondered whether they had acted wisely in keeping Kapitza at home. His dismissal from his post in 1946 and his house arrest thereafter is now attributed to his unwillingness to work on the design of nuclear weapons. Plainly Cambridge had influenced Kapitza just as he had influenced Rutherford's laboratory. But, on balance, Kapitza was well worth the trouble he caused — his Nobel Prize (for his work on superfluid helium, a consequence of what he had planned at Cambridge) and his contributions to the early stages of the Soviet space programme are proof of that. But what both East and West lost in 1934 was the embodiment that Kapitza had then become of the principle that in many circumstances science is more important than citizenship. Kapitza's visit to Britain in 1966, when his academic gown is said still to have been hanging in the Senior Common Room of Trinity College, Cambridge, came so late as simply to demonstrate how wide the gulf had become. His death is a reminder that it needs still to be bridged. □

Nuclear exports

US heads for trouble on reactors for China

Washington

THE United States is inching closer towards crowning its improving scientific and technological links with China by selling it two pressurized water reactors in the thousand-megawatt range. Senior administration officials do not expect President Reagan to clinch the deal during his visit to China this month, but they maintain that the Chinese, by making significant new commitments to the principle of non-proliferation, have removed the most serious remaining obstacle.

The US-China agreement on science and technology, signed by President Carter in 1979, has become increasingly important for the Chinese. Extended last January for another five years, it is now the most extensive intergovernmental science and technology programme involving the United States. More than 10,000 Chinese citizens are now studying or undertaking research in the United States, about half sponsored by the Chinese government; the rest study under private auspices.

An indication of the importance the Chinese attach to the agreement is their reluctance to allow disagreements with the United States, notably over Taiwan, to interfere with scientific and technical exchanges. China's biggest complaint, the imposition of stringent controls in the export of security-related US technology, has been substantially met. Last November the United States transferred China from the "special export" category to the category of "friendly but not allied" nations.

Until recently, China's failure to sign the Non-Proliferation Treaty (NPT) appeared to have placed an insuperable obstacle in the path of a nuclear agreement between the two countries. In January, however, China joined the International Atomic Energy Agency (IAEA) and Premier Zhao Ziyang, visiting Washington in the same month, emphasized China's wish to prevent proliferation.

Both the Reagan Administration and the US nuclear industry are eager to press ahead with the reactor deal. China has no operational nuclear power plants at present, and its intention to build up to ten reactors makes it the only promising foreign market, apart perhaps from Egypt, for US reactors. US manufacturers have meanwhile been prevented, while the agreement has been under discussion, from bidding for a share of the work on China's two-reactor project at Guan Dong. Their British and French competitors have not been so constrained, while Japan is supplying the pressure vessel for an indigenous 300-megawatt reactor under

construction near Shanghai.

Congress may yet stand in the way of an agreement, however. Because China already possesses nuclear weapons, its failure to sign the treaty puts no legal obstacles in the way of a sale. But the US Atomic Energy Act, as amended by the 1978 Non-Proliferation Act, contains several provisions that could offend Chinese national sensibilities and enable those in Congress who suspect China of helping Pakistan develop a nuclear weapons programme to block the agreement. The real stumbling block is a provision that allows the United States to veto the reprocessing of Chinese nuclear fuel used in a reactor supplied from the United States. Countries already thus affected complain that the provision is cumbersome, disrupting their nuclear power planning. Japan, for example, has routinely to seek US permission to ship spent fuel to France and Britain for reprocessing.

Privately, administration and congress-

sional officials concede that for a nuclear power such as China, US exertions to stress the importance of non-proliferation are symbolic rather than substantive. US policy, however, is to win international support for the Non-Proliferation Treaty by managing its nuclear relations with nuclear and non-nuclear powers on an equal basis.

In some cases, that has proved difficult. The Federation of American Scientists (FAS), for example, is concerned about the State Department's failure to enforce several agreements between the United States and the European Atomic Energy Community (Euratom) which, it believes, obliges the United States to prevent France from using its new Superphénix breeder reactor for nuclear weapons purposes. Up to 30 per cent of the plutonium to be used in the initial core of Superphénix was separated from spent-fuel originating in the United States. FAS, and several congressmen, believe the United States is legally obliged to see that none of that plutonium is used for military purposes.

France, however, refuses to identify which nuclear reactors contribute to its weapons capabilities. An attempt by the US State Department to meddle with the running of the proud Superphénix project could hardly fail to arouse some fierce Gallic hauteur.

Peter David

US nuclear power

Management blamed for failure

Washington

A MARATHON research study just completed by the Nuclear Regulatory Commission (NRC) has concluded that poor management by power companies and their contractors — not the complexity of NRC regulations — is the underlying cause of the chronic technical deficiencies of many new US reactors. The report, one of the most outspoken ever issued by NRC, says many first-time reactor builders have embraced nuclear technology with a "false sense of security" engendered by previous successes with fossil fuel plants.

NRC, which undertook the study at the insistence of Congress, looked in detail at four reactor projects (Marble Hill, Zimmer, South Texas and Diablo Canyon) that have been plagued by difficulties. In most cases, it concluded, the difficulties were due to inadequate management, not to poor craftsmanship by builders and technicians. In almost all cases, the power companies' experience had been limited to construction of fossil fuel plants, and managers approached the building of reactors without appreciating the technological challenges. NRC reports that one (unidentified) chief executive described his utility's first planned nuclear plant as "just another tea kettle".

This overconfidence, the commission found, tended to result in understaffing and the selection of contractors with

limited experience of building nuclear plants. Often, power companies used the same contractor to build the plant and to monitor its progress. In some cases, the report says, nobody was managing the project; "the project had inertia but no guidance or direction". Previous nuclear experience, NRC maintains, appears to be the single most important advantage of companies which had been able to build reactors with relatively little difficulty.

The commission concedes that it must share some of the blame for having granted construction permits to power companies that would not have received them in today's stricter regulatory environment. The study recommends that future applicants for construction permits will have to prove their overall managerial competence, both before receiving a permit and at two-year intervals thereafter. NRC is considering setting up a new advisory committee to judge the managerial ability of applicants.

The commission may also adopt a recommendation in the report to set up a system of independent inspectors to audit the progress of reactor projects every two years. NRC would establish guidelines for the audits, but the inspectors would be independent third parties whose job would be to ensure that plants were being built according to their design and licensing requirements.

Peter David

Japanese education

Nakasone plans for reform

Tokyo

THE first step towards a major reform of Japan's educational system was successfully taken last week, when Prime Minister Yasuhiro Nakasone succeeded in winning sufficient support from opposition parties to establish a reform commission.

Last month, a private advisory panel set up by Nakasone had come out in favour of drastic changes in the university and school systems. Nakasone quickly embodied these ideas in a bill setting up a special commission to recommend how best to bring about reform. With support gained from two of the smaller opposition parties, the passage of the bill in the Diet is now certain, and the commission, all of whose members will be appointed by the prime minister, will begin work on 1 July.

Both the private advisory panel and popular opinion identify the university entrance "examination hell", and its overwhelming importance in deciding the lifetime success of students, as the real villain of the educational system. Entrance into a top state university virtually assures success for life. The influential ministries and the major companies take virtually all of their staff from this elite group.

To enter an elite university requires high scores in a series of examinations which largely test students' ability at rote learning. Incredible feats of endurance are required — government surveys show that throughout their school careers, more than half of school students put in more than two hours of extra study each night, while 20 per cent are averaging three and a half hours a night. Further efforts still are required around examination time, when the dictum "Sleep five hours pass, sleep six hours fail" holds.

This pressure to succeed has warped the whole school system, down to the primary school level. Primary education has become little more than a competition for entry into a good junior high school, and junior high school a competition for entry into a good senior high school — where the cramming for the university entrance examination begins in earnest. At each stage there are tests and more tests and ever-increasing reliance on after-school cramming colleges. As Japanese schools are traditionally unstreamed and class sizes are rising towards the legal limit of 45 students, those who cannot keep up soon find themselves abandoned.

The results may include the increase in violence against teachers in junior high schools, principally from poorer students whose families have not been able to afford after-school cramming, as well as a 25 per cent a year increase in juvenile delinquency and a rapid increase in the school drop-out rate. At graduation time now, more than ten per cent of junior high schools require the presence of police guards.

What directions may reform take? One possibility raised by Nakasone's private advisory group is that individual university departments could decide their own entrance criteria: instead of the overall test score determining entry, medical schools could, for example, take the revolutionary step of interviewing candidates to see if they would make good doctors. Or college applicants could be given the choice of applying on the basis of test scores, essays or high school assessments.

Further down the system, high school entrance could be decided by taking into account teachers' reports and extracurricular activities as well as test scores. One stage of competition could even be done away with by fusing junior and senior high schools, although this would controversially require revision of the basic law on education that established a democratic education system at the end of the Second World War.

Reforms along these lines are nevertheless all popular. But another feature of Nakasone's programme explains why the Socialist and Communist parties have consistently opposed his special reform commission. Together with an "increase in flexibility" of the system, Nakasone has also called for "moral education" to strengthen the Japanese spirit of "unity, industriousness and loyalty" which he sees as stemming from the combination of Japanese "ancient national traits" with Confucianism, Buddhism and democracy. Plainly this appeal seems to the left wing a recipe for returning to the narrow dogmas of pre-war nationalism.

Teachers also worry just how such "moral" training might be implemented. Already suggested are the "retraining" of teachers, the majority of whom are members of the left-wing Japanese teachers' union which has consistently opposed Nakasone's policies, and increasing the already tight control over the contents of school textbooks. Just what might happen in this latter sphere was well illustrated two years ago when an attempt was made to reinterpret Japanese history by referring to the invasion of China as an "advance" and the forced conscription of Korean labour during the war as "labour mobilization".

Those supporting the new commission are calling for its first recommendations to be put into operation within two years. But most members of the public take a much more cynical view: proposing reform will help to boost Nakasone's popularity as he heads into the November contest for the party leadership, but the reliance on pure examination results for university entrance and the old boy network that ensures success for elite university graduates are just too deeply embedded in Japanese society for any real changes ever to be made.

Alun Anderson

UK nuclear energy

Research plans scrutinized

THE United Kingdom Atomic Energy Authority (AEA) is resisting government proposals to hive off some of its programmes and establish them as independent companies. The authority, a statutory body financed by parliamentary grant, is now the subject of a "wide-ranging overall review" by the Department of Energy. The department is looking especially hard at the work the authority does for the Central Electricity Generating Board on the safety of pressurized water reactors and advanced gas-cooled reactors, and at its non-nuclear research.

AEA officials maintain that the authority's strength lies in the wide variety of engineering specializations that it is able to bring to problems. They say that it would be weakened by the removal of any sector of its activities. Officials also point out that much of the safety research supporting the nuclear power programme is already self-financing. AEA is pressing for more freedom to initiate research into alternative methods of radioactive waste disposal; it has been a long-standing grumble that the same government department that regulates radioactive waste disposal, the Department of the Environment, is also the sole sponsor of research in the field.

The fast reactor development programme, centred at Dounreay in the north of Scotland, absorbed nearly half of AEA's net expenditure of £205 million in 1982-83. Earlier this year, the Comptroller and Auditor General criticized financial controls on the authority's spending, especially on the prototype fast reactor (PFR). The auditor's report concluded that present arrangements do not require AEA to work to sufficiently clearly defined objectives and time scales. AEA officials say the report is misleading, however, and that it appears to confuse the cost of the PFR at Dounreay and the overall research programme of which it is part.

Sir Peter Hirsch, the authority's chairman, recently explained that despite the delays which have dogged the PFR, a commercial-scale fast reactor could now be built up to 20 per cent more cheaply than would have been possible in the 1970s, when one was originally planned before electricity demand projections fell. The total cost of the fast reactor project, including work on fuel cycle development, is now expected to reach £3,700 million. Technical problems with leaking welds in the secondary coolant circuit at Dounreay have delayed the PFR but it is now expected to run at full power late this summer, and 15 per cent fuel burn-up, the target for a commercial fast reactor, is "confidently predicted". AEA insists the difficulties of the past 10 years have been overcome.

The authority is now in the process of negotiating with its foreign partners the details of the fast reactor research collaboration agreements that it signed in March. AEA has been given an assurance that the current review will not interfere with that process. UK expenditure on fast reactor development is to settle at £75 million a year for the foreseeable future (1984 prices), considerably less than the expenditure a few years ago and rather less than France and West Germany are putting into fast breeder development.

The government will now be looking for further ways of achieving savings at AEA. There are some fears that short-term financial targets may reduce the supply of nuclear engineers to a level below that necessary to support any expansion of the nuclear power programme next century. Energy strategists perceive a danger that failure to plan for future manpower needs could result in over-rapid expansion of the industry with consequent over-capacity: the French reactor construction industry is cited as an example. **Tim Beardsley**

Biological weapons

Soviet Union accused

Washington

THE US Department of Defense (DoD) said last week that the Soviet Union appears to be trying to apply genetic engineering research to an already broad biological warfare research effort that violates the Biological and Toxin Weapons Convention of 1972. Although the allegation, contained in the department's latest annual assessment of Soviet military power, is not accompanied by any evidence, the DoD pamphlet argues that genetic engineering could open up a large number of biological warfare possibilities. Natural organisms could be modified to carry diseases for which an opponent has no cure, and agents now thought too unstable for storage or biological warfare could be turned into practical weapons.

DoD says the Soviet Union has at least seven biological warfare centres, including one in the city of Sverdlovsk. It was this centre, the United States believes, which was responsible for a major anthrax leak in 1979 that may have infected as many as 3,000 people. The Soviet Union says the anthrax was caused by contaminated meat sold on the black market.

The DoD assessment reports that defence research and development by the Soviet Union is growing at a rate of six to

claims, is a coordinated campaign to poach scientific and technological expertise from the West. The Soviet Academy of Sciences and several of its institutes follow Western science and technology and subscribe to the growing number of computerized databases established in the West to disseminate findings. Although it concedes that the bulk of such technology transfer is legal, DoD says much of it is not. It complains that the Soviets acquired crucial information on magnetic bubble memory through a Hungarian scientist working at a US university on a US-funded grant.

The pamphlet says the Soviet Union has made good use of its student exchange programme with the United States. At least three-quarters of the students it sends to the United States are scientists or engineers, whereas their US counterparts tend to be from the social sciences or humanities. Soviet candidates have "nearly always" proposed research involving technology with military applications. **Peter David**

Refusnik freedom?

SOVIET Jewish scientists may have reason to hope that the regime of Mr Konstantin Chernenko will prove less harsh than that of his predecessor. Dr David Goldfarb has now been given permission to emigrate although previously he had been several times informed that, in his former work as a molecular biologist, he had had access to classified data, and therefore could not be allowed to leave the country.

Moreover, Dr Viktor Brailovskii, who has just completed a term of Siberian exile for (in effect) having hosted the Sunday seminars for "refusnik" scientists, has been allowed to return to his Moscow home. Previously prominent activists who had served terms in Siberia were not allowed to return to Moscow or Leningrad.

Even so, pessimists among the refusniks point out that only 51 Jews were allowed to leave the Soviet Union in March, the smallest number since the emigration movement started in 1968. Optimists suggest this figure was an aftermath of the Andropov era. **Vera Rich**

Hungary

Week exposes weak links

THE "Hungary Today" week in Britain (9-13 April), although it included some peripheral cultural events, was intended primarily to promote Hungarian technology. The first such event for twelve years, it included a small technical and trade exhibition in London, and briefing seminars in London and Manchester. Held in the afterglow of the visit of the British Prime Minister, Mrs Margaret Thatcher, to Hungary in February, and a successful "British week" in Budapest, despite diplomatic pleasantries, the week highlighted the many difficulties inherent in East-West cooperation.

In nuclear power, for example, Hungary is committed to a programme based on Soviet VVER light-water reactors and Soviet fuel. Hungary provides the generating equipment and the resulting station is more than a simple matching of the two components, since the Hungarians have introduced a number of improvements of their own. Notably, they have developed a special containment system for the VVER reactor which they have managed to sell back for a Soviet power station. Since it would be unrealistic to envisage British nuclear reactors in Hungarian power stations, the only possibility of major cooperation would be in the joint supply of equipment to a third (preferably neutral or nonaligned) country.

In some sectors, cooperation in knowhow seems more promising than in hardware. The Hungarian modification of the PROLOG programming system, M-PROLOG (the M stands for modular) seems to offer a number of possibilities. In one case, the construction of dry-cooling towers for conventional power stations, Hungary has to some extent stolen a march on Britain; using a principle pioneered at Rugeley in the 1960s and since abandoned in Britain, the Hungarians have secured several valuable construction contracts in arid countries. The Hungarian visitors, incidentally, seemed only too willing for future British partnership in such ventures.

But many obstacles stand in the way of further cooperation. For instance, the Csepel machine tool works can offer high-precision computer controlled machine tools at prices attractive to British entrepreneurs. But if the device breaks down, up to a fortnight of valuable production time could be lost while the maintenance engineer applies for a visa for his service call. Perhaps, in future, events aimed at promoting technological cooperation should include a session in which the technologists and customers could tell the bureaucrats how the latter could best help by removing or ameliorating tiresome non-economic constraints. **Vera Rich**



seven per cent a year and supports a network of 3,200 research institutes. The Soviet scientific enterprise, DoD ruefully points out, enjoys several advantages over that of the United States. Soviet educational institutions graduate five times as many scientists and engineers; and funding for strategically important programmes is not subject to the political uncertainties endemic in the United States.

Buttressing this indigenous effort, DoD

Laboratory animals

NIH to lay down the law*Washington*

FEDERAL rules governing the use of animals for biomedical research in the United States are to be strengthened, although a series of inspections by the National Institutes of Health (NIH) has found existing procedures to be working well. NIH have published draft rules which would make compliance with their animal welfare principles mandatory for institutions receiving NIH funds. The rules, which may be amended after a period of public comment, also provide for closer NIH monitoring of grantee institutions and for a stronger role for the animal research committees (ARCs) within universities and research institutions.

The rules were unveiled last week at an unusual conference at the National Academy of Sciences. Participants discussed the ethics of using animals in research, and humans whose lives had been saved as a result of such research gave impassioned testimony of the importance of continuing to use animals. Dr Edward Brandt, US Assistant Secretary for Health, praised the scientific community for its "responsible conduct" in the use of animals. Contemporary science, he said, is more sensitive than ever to the need for a humane approach, but this heightened sensitivity should be reflected in "stronger, more precise administrative arrangements within grantee institutions themselves".

While present NIH guidelines say research institutions seeking NIH support must express their "commitment to comply" with NIH's principles for the use of animals, the new regulations insist that institutions must regard the principles as mandatory. The regulations also provide for a more active NIH role in ensuring that the principles are indeed implemented. Institutions which choose not to seek accreditation by the American Association for Accreditation of Laboratory Animal Care (AAALAC) will have to submit annual reports describing their policy towards animals in research. They will also be subject to random inspections by Public Health Service staff.

The principles themselves have been altered little, although NIH warn that changes may follow a decision to adopt standard government-wide policies as a result of current interagency discussions. At present, NIH set out a dozen short principles covering the personnel permitted to use animals in research, the purpose and conduct of the research and the facilities in which animals are housed and transported.

Experiments involving live vertebrate animals, for example, must be performed under the supervision of a qualified scientist; and the housing, care and feeding of experimental animals must be supervised by a qualified veterinarian. The research should be "such as to yield fruitful

results for the good of society and not random or unnecessary". Experiments should be conducted to avoid "all unnecessary suffering and injury". When it does not invalidate the experiment, and when it will reduce discomfort, experimental animals are to be anaesthetized under the supervision of a senior scientist.

A significant change in existing policy, and one that may cause organizational problems, is a call for a stronger and better-defined role for ARCs. Visits to ten universities and research institutions last year apparently convinced NIH that although all institutions had set up such committees, many were "less than fully assertive" in exercising their responsibilities. As a result, the new regulations contain more specific membership requirements and set out the ARCs' role in considerable detail.

Existing rules say ARCs must have at least five members and a veterinarian. The new rules say each committee must include an individual not affiliated with the insti-

tution, a non-scientist and a practising scientist experienced in laboratory animal medicine. The committees would have extensive responsibilities. They would have oversight of an institution's general treatment of research animals, and the right to stop research projects that fail to comply with their policies. ARCs would also be expected to review all research proposals to approve the approach to animal welfare. They will not, NIH emphasize, be expected to consider scientific merit.

Few organizations have had time to assess the likely impact of the new regulations. But Ms Carol Scheman, director of federal relations for health and biomedical research at the Association of American Universities, expressed concern about the ability of ARCs to recruit members willing to deal with the large volume of work expected of them. She also warned that if the undeclared aim of the new regulations was to compel all institutions to meet the stringent standards of AAALAC, the cost could prove staggering. In 1982, NIH themselves estimated that the cost to their grantees of such an effort could exceed \$500 million.

Peter David**US science teaching****Texan textbooks evolve***Washington*

THE Texas board of education last week repealed its decade-old textbook regulations, which have required school textbooks to present evolution as only one of several explanations of the origins of mankind, and in a manner "not detrimental to other theories of origin". The rules, which have been blamed for widespread changes in biology textbooks throughout the United States, were declared unconstitutional last month by the Texas state attorney general, Mr Jim Mattox.

In a non-binding opinion, Mr Mattox said the rules had been motivated by a concern for religious sensibilities rather than for scientific truth, and therefore violated the constitutional separation of church and state. The board of education voted overwhelmingly last week to repeal the regulations. The decision means that publishers will no longer be required to print a statement on the introductory page of textbooks that any material on evolution "is presented as theory rather than fact". And it is also now easier for opponents of creationism to argue before the Texas board for the rejection of texts that have reduced or dropped references to evolution.

Pressure to repeal the Texas rules came principally from People for the American Way, a liberal pressure group committed to the separation of church and state. Like many southern states, Texas insists that all textbooks used in its schools must first be approved by the state board. Because Texas spends more than \$60 million a year on textbooks, 10 per cent of the nation's

textbook market, the rules have exerted enormous pressure on publishers.

The group claims that since the rules first went into effect, many textbook publishers have decreased or even eliminated their coverage of evolution. An analysis by Dr Gerald Skoog of Texas Tech University found that space devoted to evolution in school texts has dropped by as much as a half since the mid-1970s.

Peter David**Prison for fraud***Washington*

THREE US scientists convicted last October of falsifying the results of tests conducted under contract for the federal government are to go to prison. Announcing that he was "sending a message" to the scientific community, US District Judge John Nordberg last week imposed prison sentences on three employees of the now-defunct Industrial Bio-Test Laboratories Inc. in Chicago.

Dr Moreno Keplinger, former head of toxicology at the laboratory, was sentenced to a year in prison and four years of probation. Mr James Plank, assistant toxicology manager, and Mr Paul Wright, head of the rat toxicology section, received sentences of six months in prison and two years probation.

The three were found guilty of faking the results of experiments designed to assess the safety of an antiarthritis drug, a pesticide and an antibacterial agent used in soap (see *Nature* 303, 738; 1983). The sentences were stayed pending appeals.

Peter David

UK geology

Move from London opposed

THE British Geological Survey (BGS) is being forced to think again about a plan to move its scientific collections and library from their current prime site in South Kensington, London, to remote Keyworth in Nottinghamshire. Users protest that the move will leave the capital without a specialist geological library, while much of the library material at Keyworth will be duplicated or even triplicated.

The survey, known until this year as the Institute of Geological Sciences, is in dire financial straits. While two-thirds of its income is earned from commissioned research, it received £9 million from the Natural Environment Research Council (NERC) in 1982-83. This year, BGS's grant from NERC is likely to be cut by 50-60 per cent. The move to Keyworth is seen as a way for NERC to achieve substantial savings but is proving very unpopular

with BGS staff and users of the library and collections. Formal protests have been made by the University of London, the Palaeontographical Society and others.

BGS seems ready to accept that it will have to keep open some sort of information point in London where its publications and maps will be available for consultation. Large-scale geological maps are among the most used items in the library. But then, the argument goes, why not keep some of the books in London also, at least those that would be duplicated in Keyworth? BGS will say only that it is holding consultations. The consultations are less than urgent, however, as the buildings that will house the library and museum at Keyworth have not even been designed yet. Staff complain they are not kept adequately informed on progress.

The plan for the geological specimens is that reference collections in palaeontology and petrology will be moved to Keyworth, while those that are on public display in the existing Geological Museum in South Kensington will be transferred or loaned to the British Museum (Natural History) — BMNH — conveniently situated next door. BMNH will then eventually take over the

whole of the building now occupied by BGS.

The move of the reference collections to Keyworth is now accepted as inevitable by users, who nevertheless complain about the inaccessibility of the site (Keyworth has no railway station for example). But staff of the geological museum are very anxious that the merger with BMNH will not damage hopes of modernizing displays, some of which are hopelessly out of date (a display on oil omits to mention North Sea oil, and the museum boasts a large exhibit on ironstone mining in Northamptonshire that finished decades ago).

A NERC/BMNH working party on the merger included in its terms of reference the condition that geological sciences must be no less visible to the public in future than they have been in the past. BMNH has ambitious plans to integrate displays to cover all the natural sciences in a way neither museum can achieve alone. The Advisory Board for the Research Councils, which recommends the share-out of the science budget, approves of the proposed merger and has recommended a modest increase of resources for BMNH to take account of this. But the amount proposed (£2.3 million for each of the three years from 1984-85) will not, according to some, be enough to achieve the much-needed modernizations. **Tim Beardsley**

Rifkin bugs bug

Washington

Mr Jeremy Rifkin, the United States' leading opponent of genetic engineering and the scourge of the biotechnology industry, was back in court last week seeking an injunction to prevent the University of California, Berkeley, from going ahead with its controversial frost-retarding experiment before a court had dealt with his suit alleging that the National Institutes of Health (NIH) had approved the experiment without adequately assessing the environmental dangers.

Berkeley had decided to postpone the experiment, which would entail the first-ever release of a DNA-modified organism into the environment, when Mr Rifkin threatened six months ago to seek a restraining order. Last year, Mr Rifkin and several environmental groups filed suit in federal court challenging NIH's decision to approve the release of such organisms (see *Nature* 305, 349; 1983). That trial is not expected to begin until May, but the University of California announced earlier this month that it intended to go ahead with the experiment as soon as weather permitted.

Judge John Sirica, who is to hear the full case later this year, has not yet decided whether to grant Mr Rifkin's preliminary injunction against the university. Meanwhile, Berkeley has released a statement emphasizing its confidence in the safety of the proposed experiment, under which potato plants will be sprayed with a DNA-modified version of a bacterium which, in its natural state, enhances the formation of frost on the crops. The modified version has been stripped of its ice-nucleating properties in the hope that it will displace its natural counterpart and enable crops to survive lower temperatures. Peter David

UK universities

New blood encouraged

BRITISH universities are now in the thick of a search for 400 academics aged under 35 to join their staffs at the beginning of next academic year. This year's quota of what are called "new blood" appointments (350) is an increase of 40 per cent on the allocations first made this time last year, although the proportion in scientific subjects has decreased. The number of appointments to be made under the "information technology scheme", 46, compares with last year's total of 70 posts. Most new blood appointments will be in the physical sciences, engineering and medicine.

This year many universities were better prepared to make their applications, having had a year to rehearse their arguments. Many were caught on the hop by the timetable for applications last year. Of the 46 information technology appointments to be made this year, 35 are designed to strengthen research while the remainder are intended to provide specialist courses, an increase in the number for research.

Universities seem to have mixed feelings about the schemes. While they are glad they can make appointments they could not otherwise afford, there is resentment in some quarters that they have to justify themselves to outside bodies such as the University Grants Committee and the Science and Engineering Research Council, which assists in assessing the applications. The Committee of Vice-Chancellors and

Principals distances itself from the scheme by saying it is convinced that universities' long-term planning and management should be returned as soon as possible to the universities. Some of the appointments made earlier this year are said to have attracted outstanding scientists into university research, while in areas such as computing, where there is strong competition from industry, difficulty was experienced in filling some posts. The University Grants Committee this year warns universities not to be too fussy in their requirements.

There was also last year some confusion over the financial operation of the scheme, which lasts for only three years. At the end of three years, financial allowances for those appointed will be incorporated into universities' recurrent grant: several appointments last year were made by departments which thought the earmarked grant of up to £20,000 (£21,000 from October) would continue indefinitely. Despite the increase in the total of new blood appointments to be made this year, the total amount of money for the scheme has not been increased and the University Grants Committee is at pains to point out that there will have to be a compensating reduction in the number of appointments made next year. Many universities that were unsuccessful last year profess surprise and bafflement about the way the allocations are made. **Tim Beardsley**

Japanese science journals

Nature bites the dust

Tokyo

ONE of Japan's oldest science journals, which for the past few years has carried on its front cover both the Japanese title *Shizen* and its English translation *Nature*, is to suspend publication.

Since it was launched in 1947, *Nature* (*Shizen*) has provided a unique monthly forum for Japanese scientists from all disciplines to present their work, in Japanese, to a wide audience. The bulk of the journal is made up of a section containing articles very much in the style of *Nature*'s News and Views. The latest (and last) issue for April takes up themes ranging from grand unified theories to the twinkling of stars. Other sections provide a brief round-up of works appearing in foreign journals, many from *Nature*, and some longer and more technical articles under the heading *Ronbun* (theses).

The suspension of *Nature* (*Shizen*) by its publisher, Chuo Koron, has astonished many of Japan's scientists. Several factors seem to be responsible. In the now intense competition for the popular science readership, *Nature* (*Shizen*)'s circulation has been on the decline. Within the past three years, six new monthly science magazines, all of them illustrated with a lavishness that made *Nature* (*Shizen*)'s sober print look archaic, had acquired a combined circulation of 750,000. This "science boom", reflecting, as *Nature* had hopefully speculate (see *Nature* 305, 365; 1983), a new pride in Japan's technological achievements, seems to have been short-lived.

Circulation of the new breed of science-based magazines has slumped this year and one of them is going out of business. The lack of interest in science is even persuading the promoters of Expo '85 — the giant international science exposition due to be held in Tsukuba Science City next year — to push the sub-theme of "colour and image" rather than the basic theme of science and technology as it is "not a popular drawing card".

A further good reason for suspending publication of *Nature* (*Shizen*) came with the retirement of the editor, Akihiko Okabe, who had been running the journal for 32 years. During his long editorship Okabe has achieved something of a god-like status among Japanese scientists, and it is said that there were great difficulties in thinking of anyone who could succeed him in the post.

There is also, of course, the strange coincidence that *Nature* (*Shizen*)'s death comes just as *Nature* sets up its first Japanese office in Tokyo. Perhaps as one scientist at editor Okabe's farewell party put it, "it would be unnatural for a nation to have two natures".

Alun Anderson

Bulgaria

Computer industry goes soft

BULGARIA'S computer industry has been let down by the software sector, Ognyan Doinov, Minister of Machine Building, told the national conference of the Bulgarian Communist Party last month. Although computers similar to the Bulgarian models have elsewhere become a "powerful tool for mastering scientific and technical achievements", in Bulgaria, computers remain "dead objects" for lack of programming knowhow.

The conference was debating a new drive for "quality" in all aspects of the national economy which, Doinov said, could be achieved only by widespread modernization and automation of industry, development of indigenous technology to replace costly and often obsolescent imports and a campaign to make the whole nation computer-minded.

To begin with, training will concentrate on secondary-school and university students, whose education is already sub-

stantially job orientated. The programme will centre on Bulgaria's new "Pravets-82" personal computer, which goes into serial production this year and which, given appropriate programs, is expected to have a wide range of commercial, administrative and educational uses as well as becoming the basis for automatic quality control systems in industry.

The sociological implications of computer training have not been overlooked. Young people who fail to become familiar with computer techniques, Doinov said, will be unable to develop their talents and they will be "disappointed with both themselves and society", while those who master computer skills will be "satisfied and self-reliant". He even hinted that computers can be used for recreational purposes, thereby becoming probably the first Comecon minister to mention computer games in the august atmosphere of a party conference.

Vera Rich

Technology by bus

BRITAIN'S Minister for Information Technology, Mr Kenneth Baker, announced last week the latest in his department's seemingly unending series of special schemes to promote applications of new technology. The department, together with the Manpower Services Commission, is to spend £2.5 million on setting up a national centre to train teachers technology.

The Department of Trade and Industry sees education as an important part of its activities; for example, it was behind the project, generally regarded as a success, to provide a microcomputer in every school in Britain. Its latest scheme might seem even more obviously to be trespassing on the patch of the Department of Education and Science, but Mr Baker seems to have more success in securing finance for innovative schemes than Sir Keith Joseph, Secretary of State for Education and Science. The new centre will eventually have to be self-

financing.

The new British School Technology Centre has been developed from a teacher training project devised at Trent Polytechnic, in Nottingham. The success of the project in Bedfordshire, where it was first tried out, has persuaded the department that it would be worth promoting at a national level. Britain is very short of technology teachers and the new centre aims to educate and inspire by means of a travelling roadshow of buses equipped with all manner of miniature technological marvels. The buses will travel around Britain providing in-service training to teachers, while the new centre continues to develop its series of courses in technology education to promote to local education authorities. Many authorities are said to be showing interest, and in areas where the new courses have been used the number of pupils choosing to study technological subjects has increased dramatically.

Tim Beardsley

Brave new world — Baker's bus in action



Plant gene banks at risk

SIR — The article "Safeguarding the pool" (*Nature* 308, 109; 1984) does well to report the current international resources. In particular it highlights the move headed by a number of developing nations within the Food and Agriculture Organization (FAO) to create a "legalized framework" for genetic resources activities. This proposed agreement, yet to be ratified by member governments, is a matter of concern to those involved in plant genetic conservation.

The International Board for Plant Genetic Resources (IBPGR) has lately been described as ineffectual¹, but while the policies of IBPGR during the past decade may not be totally free from criticism, it has achieved a great deal in preserving rapidly disappearing genetic stocks on a worldwide scale². Moreover, it is now actively promoting the free exchange of conserved germplasm and supporting its utilization in plant breeding programmes throughout the world. In the long term, only plant breeding can solve the world's food production problems by producing varieties of food crops that can make the best use of irrigation and other farming inputs in developing countries. The foundation of any crop improvement programme is in turn the availability of a pool of diverse germplasm³. IBPGR has, amongst other national and international organizations, striven to make this germplasm available to all, by conserving it and distributing it freely.

One source of political controversy is the location of major gene banks in the "North" as opposed to the "South". While it is clear that the United States and the Soviet Union hold major germplasm collections, it should be stressed that these collections were established several decades ago, before the recent interest in genetic conservation. In fact, the collection held at the All-Union Institute of Plant Industry in Leningrad was established when Nicolai Vavilov, the Russian geneticist and germplasm pioneer, and his colleagues first undertook expeditions to survey genetic diversity in many crops throughout the world during the 1920s and 1930s. IBPGR has sought to redress the balance through the formation of a network of national and international gene banks.

Another political hot potato is the idea that genetic resources should be repatriated to the countries where they were collected. We believe that countries that are "gene-rich" should unquestionably have free access to their own resources. When IBPGR has sponsored collecting missions, the material collected has been divided so that a portion has remained in the donor country and a further portion has been deposited in a designated gene bank for long-term safekeeping. If genetic resources are to be repatriated with any effective purpose, then it is essential that adequate

preservation facilities and trained personnel are available to undertake the tasks at hand. Unfortunately, in many developing countries, adequate storage facilities do not exist. At the University of Birmingham we continue to train many scientists in the techniques of managing genetic resources. Most of these people have come from developing countries. However, conservation cannot often wait for such educational development. Surely it is better to have collected genetic resources before they finally disappear, and have them safely deposited in gene banks albeit in the north, than to have procrastinated while political considerations are resolved?

Some developing nations, by their current political strategy within FAO, are striving to score barely-relevant political points against the developed nations, while ignoring the urgent needs of their people. FAO may seek to create "a more institutionalized structure" for the conservationists of IBPGR. This, we are afraid will not only lead to greater "discussion" and debate, but also to inertia in conservation and irreversible loss of world plant genetic resources.

BRIAN FORD-LLOYD
MICHAEL JACKSON

*Department of Plant Biology,
University of Birmingham,
PO Box 363, Birmingham B15 2TT, UK*

1. Mooney, P.R. *Development Dialogue* 2, 1-172 (1983).
2. Hawkes, J.G. *Diversity of Crop Plants* (Harvard University Press (1983)).
3. *World Conservation Strategy* (IUCN, 1980).

Higher education

SIR — Threats to the future of higher education come from several angles and strike at different levels. Some are political. People on the right tend to look to universities primarily to improve the nation's commercial prosperity while people on the left tend to regard them as potential agents of social change. But more sinister threats arise because people fail to see the danger to the character of higher education resulting from a progressive loss of the non-material values still held, sometimes subconsciously, by a high proportion of teachers within it.

Universities and polytechnics are of course servants of the society upon which they depend, but need to recognize that this involves providing a far-sighted social critique. Some of their influence on society is indirect: they can foster a climate in which a wide variety of intellectual, scientific and aesthetic enterprises can flourish. By encouraging the idea that standards matter — in attainment; in integrity and detachment; in methods of reasoning; in the search for truth through thick and thin — they can infuse society outside with their belief that knowledge is not merely worth possessing

but authoritative when possessed. Even this could fall short of conveying a sense that man is more than a passive unit and education more than acquiring information unrelated to his fundamental needs. It may leave society with the notion that the real function of universities is to make the machine, and the machines, work better. Better for what?

An essential part of a student's experience of higher education is to belong to an environment in which the importance of sheer knowledge, and of constantly adding to it, is taken for granted. To acquire knowledge of more than facts and theories, however, demands a maintained curiosity tinged, maybe, with wonder and at times with fear. For that curiosity to become superficial can be fatal not only to the student's education but to the progress of knowledge itself.

There is still a profound and encouraging confidence alive in higher education that truth *can* be found and that things do hold together; a widespread conviction that innumerable questions are worth asking, and that answers will eventually be discovered which will not contradict one another.

ROY NIBLETT

*Pinfarthings Cottage,
Amberley, Stroud,
Gloucestershire GL5 5JJ, UK*

Multi-computer?

SIR — The recent news item on funds for computer time (*Nature* 307, 204; 1984) has an instance of art imitating nature. In multicellular organisms an assemblage of cells, suitably integrated, performs the various functions of the single, more complex cell of a large protist. Similarly, researchers at the California Institute of Technology propose using an integrated assemblage of microprocessors to provide the capabilities of a supercomputer such as CRAY. May we expect integrated multi-processors to evolve to larger, more diverse, more flexible and more powerful instruments than even the best super-computers?

STEPHEN R. COHEN

*Department of Neurochemistry,
New York State Institute for Basic
Research in Developmental Disabilities,
1050 Forest Hill Road,
Staten Island, New York 10314, USA*

Melding words

SIR — Expert though he is in the art of praising with faint damns, it ill becomes Bernard Wood (*Nature* 306, 140; 1983) to protest at the creation of new words without checking his dictionary. Clearly, he does not play cards.

JOHN GRIBBIN

*39 Southover High Street,
Lewes,
East Sussex BN7 1HX, UK*

Publish abstracts

SIR — J.R. Metcalf (*Nature* 15 March, p.222) asks whether abstracts should be included in hard copy and computerized databases. If something is published or recorded, whatever its merit or scientific value, it is essential that the bibliographic control exists to identify and find it.

I, for one (as a librarian and information scientist), would totally support, and indeed use, any secondary source of information that included the coverage of abstracts of this nature, as well as full papers appearing in conference proceedings.

SHEILA M. MOULD

*MRC Clinical and Population
Cytogenetics Unit,
Western General Hospital,
Edinburgh EH4 2XU, UK*

SIR — Many databases such as *Biological Abstracts* include abstracts of papers presented at conferences. The range of quality of these abstracts (as indeed with full papers) is large, but as a database user in the pharmaceutical industry, I have found many occasions when the first announcement of a new drug is through such publication. Consequently I fully support their inclusion in computerized databases.

In *BIOSIS (Biological Abstracts)* it is a simple matter to exclude all such abstracts, if desired, since they appear in the "Register of Research Methods" section of that database. In other databases it is often easy to identify such abstracts because they are normally only a single page in length, or they may appear in a proceedings supplement with annotated pagination. A competent database searcher can therefore avoid inundating the researcher with such material if requested.

Surely on balance it is better to have such abstracts included and sift out the unwanted material than to proceed with research in ignorance of related and potentially important information.

I. MCFARLANE

*Organon Laboratories Limited,
Newhouse,
Lanarkshire ML1 5SH, UK*

No simple answer

SIR — Jeremy Bray (*Nature* 1 March, p.9) says that "a reduction of salt, sugar and saturated fatty acids and an increase of dietary fibre in the diet of the nation would substantially reduce mortality and improve health in the middle years of life". This talismanic theory was propounded a few years ago by Senator George McGovern. It is an illustration of "Mencken's law", which states "For every complex problem, there is a simple solution, and it is wrong". Nations do not eat diets, individuals eat diets. You cannot determine individual intake by dividing total consumption by number of people. Too much salt, sugar or fats, or even too much water, are unhealthy for an individual, but salt, sugar and fats

are basic nutrients. There is debate as to whether excesses of unsaturated fatty acids are actually more harmful than are excesses of saturated fatty acids. "Fibre" is a variable entity. It is sometimes measured as "crude fibre" by weighing residues of foodstuffs after extracting with ether, dilute acid and dilute alkali¹. It is also described as a mixture principally of cellulose, hemicellulose, lignin, pectin and gums that is not digested in the intestinal tract². "Fibre" is now so fashionable that the addition of sawdust to bread in the United States was recently proposed, which would be great for termites.

Non-smokers and non-drinkers who eat lots of meat and sugar are said to live longer than the average person³. Salt intake needs to be reduced in susceptible individuals who consume excessive quantities, rather than "in the diet of the nation". Food should be eaten in variety and in moderation, and enjoyed. THOMAS H. JUKES
*Department of Nutritional Sciences,
University of California, Berkeley,
Oakland, California 94608, USA*

1. *Composition of Foods, Agriculture Handbook No.8* (USDA, Washington, 1963).
2. Story, J.A. & Kritchevsky, D. in *Biochemistry of Nutrition*, 189-206 (University Park Press, Baltimore, 1979).
3. Herbert, V. *National Forum*, Phi Kappa Phi Journal, Winter 1984, pp.3-7.

The fly domesticated

SIR — Undoubtedly the method of F.E.G. Cox¹, permitting flies to be trapped (rather than swatted) is kind. Yet the problem of what to do with the captured fly remains. Since the intent is to avoid killing the fly, only two options present themselves: (1) release the fly out of doors or (2) contain it in a "flyhouse" in part of the home or laboratory. The second option is lamentably untenable in view of the fact that the flies thus adopted would, given adequate nourishment, soon yield a host of young, requiring frequent expansion of their domicile. The first option is therefore the only acceptable course but can result in tragedy if practised during the winter season when outdoor temperatures frequently drop below freezing. In such a climate, releasing the fly out of doors would, in a high percentage of cases, lead to its death due to exposure to the cold² (flies are poikilothermic).

A partially acceptable solution to the dilemma was suggested to me by A.S. Glass: if the fly is released at the boundary between indoors and out, at the window sill by an opened window, for example, the fly may leave the room spontaneously. If, on the other hand, it returns to the room it may be recaptured and again released, the process being repeated indefinitely (with suitable rest periods between trials) or until satisfaction is achieved. Such "encouraged suicide" was long accepted, in order to permit the survival of some members of the group, among some³ although not all⁴ traditional human societies living in correspondingly harsh climates.

Interestingly, nobody^{1, 5-7} has yet suggested the possibility of coexistence. The fly is a most intriguing creature. It is small and requires virtually no care, yet will literally eat from one's hand. It thus has many of the attributes of a fine house pet.

KENNETH L. BELL

*Department of Chemistry,
New York University,
4 Washington Place,
New York, New York 10003, USA*

1. Cox, F.E.G. *Nature* 307, 8 (1984).
2. Dethier, V.G. *To Know a Fly*, 23 (Holden-Day, New York, 1962).
3. Freuchen, P. *Book of the Eskimos*, 194-206 (World Publishing Co., 1961).
4. Vanstone, J.W. *Point Hope, an Eskimo Village in Transition* (University of Washington Press, 1962).
5. Gray, E.G. *Nature* 304, 678 (1983).
6. Miller, A. *Nature* 305, 664 (1983).
7. Kruthoff, P.N. *Nature* 305, 570 (1983).

Count is OK

SIR — R. F. Entwistle (*Nature* 308, 222; 1984) is supporting a lost cause in opposing the use of the "count" as a pseudo-unit. Language changes with use, and radioactivity "counts" made the transition from "lab-speak" to the respectability of regular appearance on the pages of biochemical literature at least 30 years ago.

For a long time this editorial office tried to persuade authors writing for its journals that "to fast", as an intransitive verb, was incorrectly used in expressions such as "fasted rats". The latest supplement to the *Oxford English Dictionary* reveals the influence of biologists on grammar; "to fast" now receives recognition as a transitive verb, with the usage sanctioned by quotations from your own distinguished pages.

TONY EVANS

*Editorial Office,
The Biochemical Society,
7 Warwick Court,
London WC1R 5DP, UK*

Pagiarism claim

SIR — It is well known that plagiarism or appropriation of ideas or passages from another work or author exists in the scientific community today. Another, contrasting practice exercised by an alarming proportion of authors submitting papers today is the wilful omission of references to previous work in the specific subject. By doing so not only do they pretend that their work or idea/s is/are original but also deceive the referees and the peers who, unfortunately, may be unable to keep up with the explosive growth of information which we witness today. As there is no specific word in the English language to describe this practice I wish to suggest the word *PAGIARISM* to denote Purposeful Avoidance of making references to past literature or ideas with intention of deceiving the community, at large.

LAKSHMAN P. SAMARANAYAKE
*Glasgow Dental Hospital and School,
378 Sauchiehall Street,
Glasgow G2 3JZ, UK*

Extinctions by catastrophe?

The suggestion that the disappearance of species from the surface of the Earth has been periodic has stimulated contradictory explanations and also a diversity of ways of dealing with information.

DOES the periodic extinction of species on the surface of the Earth arise because the Sun has a previously unknown dwarf companion which disturbs the Solar System's cometary cloud, or because the Solar System as a whole oscillates above and below the plane of the Galaxy? And, in any case, how solid is the assertion that during the past 250 million years, there has been a statistical regularity in the timing of the disappearance of large numbers of species from the surface of the Earth?

These questions are certain to be provoked by the group of five articles appearing on pages 709 to 720 of this issue. Professor Anthony Hallam on page 686 gives some of the reasons why open-mindedness on the last question is still appropriate. Readers who are perplexed by the stark contradiction between the two explanations now advanced should not, however, blame the authors of the two hypotheses but, rather, the conventions, acceptable and otherwise, which regulate, and sometimes fail to regulate, such developments.

One obvious difficulty is that all five of the articles now published refer in one way or another to an article by David M. Raup and J. John Sepkoski, of the University of Chicago, which appeared only in the February issue of the *Proceedings of the National Academy of Sciences* (81, 801-805; 1984). That publication is itself the culmination of a long process — a detailed compilation by Sepkoski (*Milwaukee Contr. Biol. Geol.* 51, 125; 1982) of what is known about the dates of the extinction of species on the surface of the Earth, partial consideration by one or other of the authors of the significance of this mass of data and eventually their conviction that there is probably a periodicity of 26 million years in the occurrence of extinctions. This conclusion seems to have been formed in their minds by May last year, when one of the authors described the result at one of the Dahlem conferences in Berlin. The article in which the conclusion is embodied was submitted for publication the following October, and has now appeared in print.

There is nothing remarkable about this sequence of events. Raup and Sepkoski are not the first to have suggested that there may be a periodicity in the pattern of species extinction, and have been properly cautious in putting forward their new interpretation of the data. The problem is notoriously difficult. The fossil record is

so poor that it is not possible to tell when individual species became extinct. It is necessary instead to work with taxonomic families, but even then it is possible to place their extinction only somewhere within the confines of a stratigraphical stage, of which 39, with different durations, are placed by international convention within the past 250 million years.

One obvious statistical pitfall is that the conventionally accepted pattern of stratigraphical stages might itself conceal a periodicity. Raup and Sepkoski break new ground in the analysis of data-sets thought to harbour periodicity by their use of computer-based tests of the reality of their inferred period of 26 million years. The authors themselves acknowledge that the conclusion that species extinction is a periodic phenomenon will be jeopardized if the stratigraphical time scale proves to be incorrect (Hallam, *op. cit.*).

So why have Raup and Sepkoski prevailed when others advocating periodic extinctions have done less well? Obviously it has helped that their statistical analysis has been done in a way that shows them to be fully aware of the pitfalls. But it is proper to acknowledge that the intellectual climate has changed in favour of catastrophism, especially now that Luiz and Walter Alvarez appear to have proved their original case that the massive extinction at the end of the Cretaceous period was caused by the impact of some extraterrestrial object (see *Science* 223, 1174-1186; 1983). Indeed, Raup and Sepkoski themselves say at the end of their article that a biological explanation of their periodicity "seems incredible" and that they favour "extraterrestrial causes".

Extraterrestrial explanations they have now stimulated in abundance, partly because preprints of their article appear to have been freely circulated in advance of publication. What seems now to have happened is that the statement by two careful investigators that extraterrestrial phenomena may be responsible for mass species extinctions has been taken by the astrophysicists as a signal to put forward models to account for what may have occurred.

Two important issues arise from this, the first of them concerned with the process of professional communication and, in particular, with the great confusion that may be caused by the widespread circulation of preprints. This practice, usually intended as a courtesy to colleagues elsewhere (and

welcomed by the recipients except when the effect is to pre-empt a patch of intellectual territory), can also be thoroughly unhelpful to other people. The most obvious complaint against the system is that it is discriminatory, excluding from the circle of those in the know people who happen not to be on the authors' mailing list. On this occasion, the first accounts of models to explain periodic extinction were so swiftly on their way to *Nature* that, if revision of their content had not been recommended, it would have been possible for them to have appeared in print before Raup and Sepkoski.

Confusion has been further augmented by the circumstance that a full account of the article by Davis *et al.* (see p.715) appeared in the *San Francisco Chronicle* in mid-February, and that a workshop on the subject (organized by Dr Luiz Alvarez) was held at Berkeley, California on 3 and 4 March, and has since been agnostically reported (*Science* 223, 2178; 1984). (A representative of *Nature* was invited but could not attend, perhaps, in the circumstances, fortunately.) One result of these developments is that criticisms of some aspects of the group of papers now published are already flowing in.

Developments such as these constitute a kind of nonsense. In the normal course of events, it is entirely proper and indeed essential that people should talk about their discoveries at conferences, and give their colleagues advance notice of what they plan to publish, while the occasional publication of the details of some new discovery in the general press does not spell the end of the world. But the concatenation of all these events can create serious doubts in people's minds about the proper attribution of novel ideas to the participants.

The second issue raised by the two groups of articles now published is more interesting because it draws attention to the difference of what may be called style that appears to distinguish astrophysicists from, say, biologists. The models now proposed for periodic extinctions are lent credibility only because Raup and Sepkoski have put forward a conclusion that cries out for an explanation. Reflection will show that astrophysicists often have no other way in which to make progress. The fact that, on this occasion, two contrasting models seem to be equally plausible must, for the time being, be counted an unlucky accident.

John Maddox

Evolution

The causes of mass extinctions

from A. Hallam

THE traditional darwinian view has it that organisms evolve and become extinct primarily as a result of competitive interactions, with changes in the physical environment being of subordinate importance. It has, however, become increasingly apparent in the last few years that organic turnover through time is characterized by long periods of relative stability punctuated by geologically brief episodes of mass extinction, during which a significant proportion of the Earth's biota is killed off. Evolution is seen to have had a substantial opportunistic component, with survivors radiating into the ecological niches vacated by extinction. What caused the mass extinction events remains a contentious issue of enduring fascination, as witnessed by a clutch of papers in this issue of *Nature*, inspired by the recently published claim of Raup and Sepkoski (*Proc. natn. Acad. Sci. U.S.A.* **81**, 801; 1984) to have detected a 26-million-year periodicity in the mass extinctions.

Palaeontologists have generally looked for an interpretation of mass extinctions in terms of changing geography or climate. Although as long ago as 1950 Schindewolf suggested that the spectacular end-Palaeozoic extinction event was caused by an increase in cosmic radiation and a number of other extraterrestrial hypotheses were put forward in the following years, in the absence of any supporting evidence none was taken very seriously. The situation changed radically in 1980 when Alvarez *et al.* (*Science* **208**, 1095) reported an abnormal enrichment in iridium and other noble metals in clay layers at the Cretaceous-Tertiary boundary, and interpreted it as a by-product of an asteroid impact which caused an expulsion of terrestrial dust on such a scale that sunlight was blotted out for several weeks, causing the simultaneous extinction of plankton and dinosaurs and many other groups. The Alvarez paper generated great excitement and has already provoked a huge literature.

The matter has been taken a stage further by Raup and Sepkoski. Using Sepkoski's compilation of the temporal ranges of marine vertebrate, invertebrate and protozoan families and a combination of Fourier analysis and Monte Carlo simulation, they claim to have detected a statistically highly significant 26-Myr cyclicity in mass extinction events since the late Permian (~250 Myr). Though they do not commit themselves to any particular interpretation they favour an extraterrestrial cause because of the regularity of the cyclicity. By circulating a preprint of their paper they, in effect, invited astronomers to propose explanatory models. The astron-

omers have responded with alacrity and devote five papers in this issue to the subject.

Davis *et al.* (*Nature* **308**, 715; 1984) and Whitmire and Jackson (*Nature* **308**, 713 1984) have independently put forward a model that postulates the existence of an unseen companion star to the Sun, occupying a highly eccentric orbit. When near the perihelion it is brought into the dense inner region of a comet cloud and by perturbing the cometary orbits initiates an intense comet shower, leading to a series of terrestrial impacts lasting up to a million years. Schwartz and James (*Nature* **308**, 712; 1984) and Rampino and Stothers (*Nature* **308**, 709; 1984) instead propose galactic models. The former authors speculate that long-term changes in cosmic radiation flux due to the Sun's oscillation about the galactic plane have provoked sufficient alterations of the biosphere to cause mass extinctions. Rampino and Stothers' reanalysis of Raup and Sepkoski's data leads them to suggest an extinction cyclicity of ~30 Myr, which correlates strongly with a galactic cycle. Extinctions might have been caused when the Earth passed through interstellar gas or dust clouds, or intercepted a cometary shower. Alvarez and Muller (*Nature* **308**, 718; 1984) have subjected the records of large impact craters to time-series analysis and conclude that most occur on a 28.4-Myr cycle. Because only 13 craters met their rigorous criteria, however, their conclusion must be viewed with caution.

In assessing the value of these speculative papers it is clearly necessary first to scrutinize the Raup and Sepkoski analysis on which they are based. The accuracy of the geological time scale used is obviously crucial if a convincing regular cyclicity is to be demonstrated. Of two recently published time scales, the Harland and the Odin, Raup and Sepkoski prefer the former without giving any reason beyond the fact that it is extremely similar to the Geological Society of America time scale. That is unsurprising since the compilers of both time scales uncritically used the same data (of Armstrong). As Raup and Sepkoski acknowledge, there are important differences between the Harland and Odin time scales, with only moderate correlation ($r = 0.48$), and there are further differences between these and a new Geological Society of London time scale to be published shortly. The problem is that before the mid-Cretaceous, the accuracy of Mesozoic dating is generally rather poor, with an error in the determination of stage boundaries of 5 Myr or more.

The 26-Myr cyclicity can only be valid if, despite the difficulties referred to above, the Harland time scale is the most accurate.

Even then there would be problems. Thus Raup and Sepkoski predict on their cyclicity model a late Norian extinction event at 221 Myr, only 2 Myr earlier than the Harland date for the end of the Norian. But the Harland time scale assumes equal time spans of 6 Myr for the Norian and succeeding Triassic stage, the Rhaetian. This cannot be accurate, because the Norian has seven ammonite zones and the Rhaetian only one. Many authorities therefore believe the Rhaetian should be relegated to a substage of the Norian or abandoned altogether. Thus the late Norian extinction probably took place only 1–2 Myr before the end of the Triassic, dated as 213 Myr by Harland, 205 Myr by Odin and the Geological Society of London and 208 Myr by Armstrong (in the Odin volume), who has revised his older estimate as used by Harland.

Of twelve extinction events recognized by Raup and Sepkoski, the mid-Miocene and Olenekian ones are admitted to be dubious and should be eliminated from a rigorous analysis. The Bajocian, Callovian and Hauterivian events are minor and deviate by 6–8 Myr from the predicted cyclicity. Since the most precisely dated event, the end-Cretaceous, is used as the reference, this leaves only five events — the late Permian, Pliensbachian, Tithonian, Cenomanian and late Eocene — where the disparity with the Harland time scale is no more than 1 Myr. Bearing in mind the inaccuracy of stage-boundary dating (in the Odin time scale the end of the Tithonian is dated as 14 Myr younger than in the Harland time scale), the evidence hardly seems to warrant Raup and Sepkoski's bold conclusion that "It seems inescapable that the post-late Permian extinction record contains a 26-Myr periodicity".

Whereas the only evidence for an extraterrestrial cause of mass extinctions is the iridium anomaly at the Cretaceous-Tertiary boundary — and even that remains controversial — there is a wealth of evidence relating mass extinctions to events on Earth. As discussed at some length in my forthcoming article (*A. Rev. Earth planet. Sci.* **12**, 205; 1984), the most obvious relationship is with sea-level changes, which have exhibited a cyclicity on more than one scale throughout the Phanerozoic. In several papers in the 1960s, Newell proposed that mass extinctions of marine groups took place as a result of the drastic reduction in area and quality of habitat when extensive epicontinental seas regressed because of a fall in the sea level. This seems to me to point the way to the best general explanation, but the truth is clearly more complicated. In some cases, as in the end-Pliensbachian and end-Cenomanian events recorded by Raup and Sepkoski, and also quite possibly in the end-Triassic, late Devonian and end-Ordovician events, it might have been anoxia (as recorded by the extensive black shales deposited during the immediately succeeding rise in sea level) that caused the critical deterioration of the environment; it

is noteworthy that the Cenomanian, Pliensbachian and Devonian events were not preceded by significant falls in sea level. In other instances a change in climate is implicated. Thus, although the fall in sea level at the end of the Permian, Triassic, Cretaceous and Eocene would have extended the habitat area of large terrestrial vertebrates, they suffered mass extinctions perhaps because of an increased continentality of the terrestrial climate, with greater seasonal temperature extremes. There is good evidence that global air and seawater temperatures were falling well before the end of the Cretaceous and Eocene, and the fall in sea level at the end of the latter period may well have been linked with the initiation of the Antarctic ice cap.

Many problems and uncertainties remain, and there is much to be learned about the response of particular groups of organisms and whole ecosystems to changes in the physical environment; the plankton extinctions at the end of the Cretaceous are still an enigma. The point is that these problems are readily amenable to analysis, ideally by teams of palaeontologists working in conjunction with stratigraphers, sedimentologists and geochemists. Before astronomers indulge in further speculations about the cause of mass extinctions they would do well to learn something about the rich stratigraphical record of their own planet. □

A. Hallam is Lapworth Professor of Geology at the University of Birmingham, PO Box 363, Birmingham B15 2TT.

Gene regulation

Latterday lessons of lambda and lac

from Geoffrey North

WHEN it comes to understanding the recognition by proteins of specific DNA sequences and the control of gene expression, molecular biologists can still learn a lot from those classical prokaryotic systems, bacteriophage λ and the *lac* operon. So much was clear at the end of a recent meeting* which demonstrated that molecular geneticists should not be too hasty in ascribing novel mechanisms to eukaryotic regulatory sequences, such as 'enhancers', which at first sight appear quite distinct from anything seen in prokaryotes. Furthermore, it was revealed that we are on the verge of the first high-resolution structure of a complex between a regulatory protein and DNA.

DNA-protein recognition

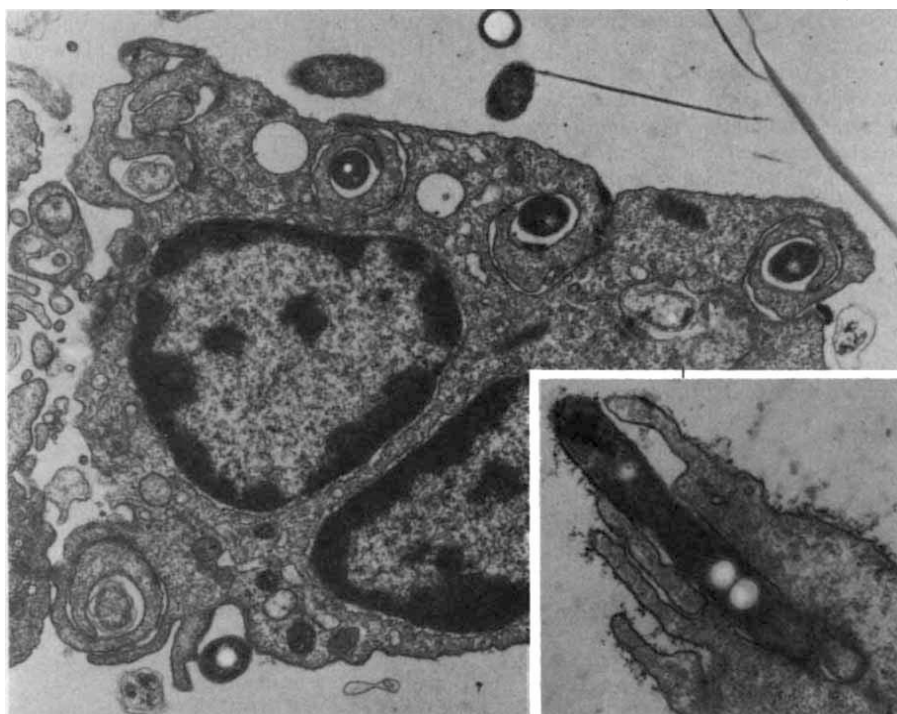
When the structures of the catabolite gene activator protein (CAP)¹ of *Escherichia coli* and the *ci*² and *cro*³ repressors of bacteriophage λ were solved by X-ray crystallography a few years ago, they immediately suggested models for how the proteins might bind to DNA and recognize specific sequences. In spite of a brief diversion in the case of CAP into binding to left-handed DNA¹ (shown to be incorrect on

topological grounds), all three proteins seemed to have arrived at a similar solution to the problem of recognizing particular linear arrangements of base pairs tucked away on the inside of a fairly uniform DNA helix. Each protein contains a pair of α -helices, one of which fits neatly into the major groove of an essentially B-form helix, where contacts are made between exposed functional groups on the bases and amino acid side chains¹⁻⁴. And as M. Ptashne (Harvard University) reported in Cologne, simply by swapping the α -helices of the *ci* and *cro* repressors of the lamboid bacteriophage 434 that were predicted to fit into the major groove, it has now been confirmed that they confer the sequence specificity of DNA binding.

The models make detailed predictions about the importance of particular amino acid-base pair interactions, and several speakers reported how these predictions are borne out by recent experiments. One approach has been to use specifically altered binding sites on the DNA to select for mutant regulatory proteins with altered sequence specificity. P. Cossart (Institut Pasteur) described how two symmetrically related single-base changes in the CAP site of the *lac* operon both selected for alterations to residue Glu 181 of CAP. This is consistent with the accepted model for dimeric CAP binding to a site of approximate inverted sequence symmetry, and fits precisely with models of sequence recognition by CAP (Cossart and T. Steitz, Yale University).

The important conclusion drawn by R. Sauer (MIT) on the basis of an analysis of mutations that increase or decrease the affinity of repressor proteins for their binding sites (and also some that actually alter the specificity of binding in a defined fashion⁵) is that the presence or absence of a single interaction between a base and an amino acid can alter the affinity of protein binding by three orders of magnitude. This explains how the specificity for binding of the appropriate repressor to the *Gal* and *lac* operons is conferred by a single-base-pair difference (B. Müller-Hill, University of Cologne).

This detailed picture of how prokaryotic gene-regulatory proteins recognize DNA of specific sequence should be completed in the near future by the X-ray crystallographic determination of the structure of protein-DNA complexes. S. Harrison (Harvard University) reported that although the diffraction pattern obtained for



A HUMAN monocyte being invaded by several *Legionella pneumophila* — the bacterium that causes Legionnaires' disease. Phagocytosis of the bacteria by monocytes and by other phagocytic cells involves a novel process whereby cell pseudopods enclose individual bacteria, most of which contain a translucent fat vacuole. A section along the axis of a coil (inset) shows how a bacterium is trapped between cell projections. Once phagocytosed, *L. pneumophila* bacteria multiply within the cells. From M.A. Horwitz *Cell* 36, 27 (1984).

*The Cologne 1984 spring meeting on molecular biology was held at the University of Cologne, 7-9 March 1984.

cocrystals of the repressor of bacteriophage 434 and a 14-base-pair (bp) synthetic operator has not yet been converted into a molecular structure, it is already clear that the DNA does not deviate noticeably from the B-form⁶.

Positive control

When the λ cI repressor binds to its specific recognition sequences — the tripartite left and right operators of the λ genome — it not only turns some genes off but also turns others on⁷. In doing so cI opposes the actions of the cro repressor in a way that mediates the switch between lytic and lysogenic growth states of the bacteriophage. The proximity of the binding sites for cI to the promoters that it activates suggests that it interacts directly with RNA polymerase. The suggestion is borne out by the existence of mutant cI proteins, no longer capable of positive control, whose amino acid alterations are localized on the surface of the protein just where contact with RNA polymerase would be expected⁸ and by the fact that these cI mutations can be complemented by second-site alterations to RNA polymerase (Ptashne).

CAP activates a number of operons, including *lac*, when intracellular cyclic AMP levels are elevated in the absence of catabolite repression. The way CAP works is of particular interest because, like the enhancer sequences recently discovered in eukaryotic cells (see below), some CAP sites are spatially separate (up to 20 bp upstream) from the promoters they regulate. Although this raises the possibility of control by some distant effect on the promoter, perhaps by distortion of the DNA helix, two results reported at the meeting favour a cI-like mechanism of direct interaction between CAP and RNA polymerase: first, the RNA polymerase selected to complement defective cI proteins also interacts more strongly with CAP at the *lac* operon (Ptashne); second, single-base-pair insertions and deletions between a CAP site and a promoter seem to disrupt the ability of CAP to activate the promoter, which is inconsistent with simple models of action-at-a-distance (M. Caruthers, University of Colorado).

In the yeast *Saccharomyces cerevisiae*, a number of genes have been found to be regulated by sequences with properties intermediate between those of CAP sites and the enhancer sequences of multicellular eukaryotes. Like enhancers, these 'upstream activation sites' (UASs) seem able to influence a gene within a range of distances from its promoter⁹, though as yet they have not been shown to work when placed downstream of a gene, as enhancers can.

So are UASs recognized by proteins that, as CAP seems to, interact directly with RNA polymerase? Preliminary results reported by Ptashne suggest they may be.

One clue comes from experiments on an UAS that regulates a divergently trans-

cribed pair of genes involved in galactose metabolism. The genes are normally induced by galactose, which stimulates the binding of the *Gal4* gene product to the UAS, perhaps thereby creating an efficient entry site for RNA polymerase. When the UAS is experimentally moved close to the promoter of one of the genes, the gene becomes partially constitutively active, perhaps because the UAS has some activity on its own as an entry site if it is near enough to the promoter to minimize the chance that the RNA polymerase will drop off before initiating transcription. Another clue is that when binding sites for the bacterial repressor protein Lex A are inserted between the UAS and a promoter, activation of the gene is prevented by Lex A, which is consistent with a model in which RNA polymerase enters at the UAS and 'slides' to the promoter.

On to metazoans

The fashionable 'enhancers' are eukaryotic examples of DNA sites that are distant from gene promoters and yet can influence their activity. Enhancers can work at distances of up to several kilobases from a gene, and even downstream of the transcription unit¹⁰. These properties have led to a number of suggestions as to how enhancers may work, but if, as intimated above, the mechanism is analogous to those postulated for CAP sites and UASs, then the simplest of these suggestions — that they are entry sites for RNA polymerase — may turn out to have been the right one.

There is, however, another important property of metazoan enhancers not applicable in bacteria or yeast: they are crucial to the peculiarly metazoan problem of cell-specific gene expression^{11,12}. The clearest examples of this are the enhancers located between the J and C exons of immunoglobulin heavy and κ light-chain genes¹³⁻¹⁷ which confer on a gene to which they are linked the ability to be expressed in B lymphocytes — but not in any other cell type. And in Cologne, P. Gruss (EMBL, Heidelberg) reported that the recently discovered lymphotropic papovavirus carries an enhancer active in lymphocytes of both the B and T lineages.

Effort is now being concentrated on identifying the cellular factors presumed to interact with enhancers. As a first step, Gruss reported that co-transfected enhancer sequences can be demonstrated to compete for some cellular component essential for their activity¹⁸. And W. Gilbert (Harvard University) mentioned a preliminary result, obtained with a potentially powerful new technique¹⁹, which might have provided the first glimpse of the factor in B cells that recognizes the immunoglobulin heavy-chain gene enhancer. The technique involves the direct sequencing, by the Maxam and Gilbert method, of regions of genomic DNA picked out by very highly labelled probes. Direct sequencing allows the

detection of all methylated base pairs — not, as before, just those recognized by isoschizomer restriction enzymes — and thus, for the first time, makes it possible to explore the importance in gene regulation of methylation of any specific DNA sequence. At the same time, protein bound to DNA can be detected by its ability to protect specific bases from chemical modification *in vitro*. The signs are that the pattern of methylation protection at the immunoglobulin heavy-chain gene enhancer is different in B cells from in a cell type in which it is inactive — presumably because of protein bound to the enhancer in B cells alone.

Enhancers may not be the only sequences conferring tissue specificity on the promoters of genes in multicellular eukaryotes: R. Tjian (University of California, Berkeley) described the identification of a protein (Sp1) which not only confers specificity upon the simian virus 40 (SV40) early promoter by binding to the 21-bp repeat sequences that constitute the SV40 early promoter itself²⁰, but also binds to the promoter of a cellular gene, identified by probing a monkey genome library with SV40 early-region DNA, and turns it on. So it looks as though SV40 uses a promoter representative of a class of cellular promoters — presumably ones active in a fairly wide range of cell types.

Even when the proteins that regulate the expression of genes in multicellular eukaryotes have been identified and the way in which they work is understood, the question will remain as to how they come to be expressed in some cells and not others. Attention will then focus on what regulates the expression of the regulators. In the case of λ the answer is that the regulatory proteins regulate their own expression; could it be that the recent suggestion that the cellular *myc* gene regulates its own expression²¹ is telling us that here again the lessons of lambda and *lac* should not be forgotten. □

- McKay, D. B. & Steitz, T. *Nature* **290**, 744 (1981).
- Pabo, C. O. & Lewis, M. *Nature* **298**, 443 (1982).
- Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. *Nature* **290**, 754 (1981).
- Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthews, B. W. *Nature* **298**, 718 (1982).
- Youdarian, P., Vershon, A., Bouvier, S., Sauer, R. T. & Susskind, M. M. *Cell* **35**, 777 (1983).
- Anderson, J., Ptashne, M. & Harrison, S. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1307 (1984).
- Johnson, A. D. *et al. Nature* **294**, 217 (1981).
- Hochschild, A., Irwin, N. & Ptashne, M. *Cell* **32**, 319 (1983).
- Guarente, L., Lalonde, B., Gifford, P. & Alani, E. *Cell* **36**, 503 (1984).
- Yaniv, M. *Nature* **297**, 17 (1982).
- Boss, M. A. *Nature* **303**, 281 (1983).
- Hogan, B. *Nature* **306**, 313 (1983).
- Hayday, A. C. *et al. Nature* **307**, 374 (1984).
- Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* **33**, 719 (1983).
- Banerji, J., Olson, L. & Schaffner, W. *Cell* **33**, 729 (1983).
- Neuberger, M. *EMBO J.* **2**, 1373 (1983).
- Mills, F. C. *et al. Nature* **306**, 809 (1983).
- Schöler, H. R. & Gruss, P. *Cell* **36**, 403 (1984).
- Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991 (1984).
- Dynan, W. S. & Tjian, R. *Cell* **35**, 79 (1983).
- Leder, P. *et al. Science* **222**, 765 (1983).

Geoffrey North is an assistant editor of *Nature*.

Nuclear structure

Bending over backwards

from Neil Rowley

IRREGULARITIES in the rotational behaviour of deformed heavy nuclei can provide us with valuable information on their internal dynamics. The most striking irregularity is 'backbending', where the rotational frequency is seen to drop despite an increase in angular momentum. Although this effect has frequently been observed for many rare-earth nuclei, only with the recent paper of Spreng *et al.* (*Phys. Rev. Lett.* **51**, 1522; 1983) has backbending been established for a nucleus (^{244}Pu) in the actinide region.

Figure 1 schematically shows the moment of inertia $\mathcal{J}$ of a 'backbender' as a function of its rotational frequency ω (deduced from the frequencies of the emitted γ rays) with the numbers along the curve indicating corresponding values of the nuclear spin I (in units of $\hbar$). A nucleus may have many states of a given spin but those which display this peculiar behaviour are the so-called 'yrast' states of the system, the states of highest spin for a given excitation energy which therefore display the most highly collective rotational behaviour. (The yrast states of ^{244}Pu , up to $I=26$, are populated by Coulomb excitation with a beam of energetic ^{208}Pb ions).

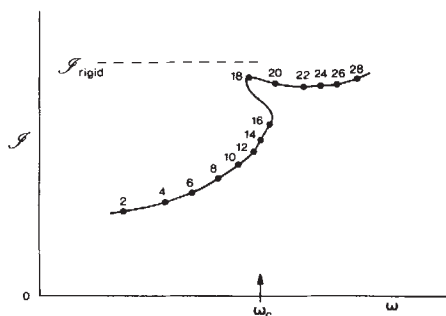


Fig. 1 Moment of inertia $\mathcal{J}$ of a backbending nucleus as a function of rotational frequency ω . The numbers along the curve indicate the spin I in units of $\hbar$. The nucleus 'backbends' at the 'crossing' frequency ω_c (typically about 10^{20} per second).

Although a sudden increase in spin and a large jump in the moment of inertia, usually to a value close to that of an equivalent rigid body, is associated with backbending, many heavy nuclei display such increases in $\mathcal{J}$ without the curve bending back on itself but merely 'upbending'. Both types of behaviour have been observed for rare-earth nuclei but only gentle upbending had been seen in the actinides until the report of Spreng *et al.* Indeed, nuclei of a very similar mass and charge (for example ^{238}U , ^{248}Cm) show no sign of backbending. Thus, although these heavy nuclei contain a large number of nucleons, their collective behaviour is significantly

modified by the removal or addition of a few neutrons or protons. It is this selectivity which makes the nucleus such an intriguing many-body system.

In order to understand the above phenomena one must consider the various forces acting on a nucleon bound to a deformed rotating nucleus. Essentially, these are, first, the average force due to all the other particles (described by a deformed potential); second, short-range 'residual' interactions arising from close collisions between pairs of nucleons; and third, the inertial forces (centrifugal and Coriolis) stemming from the rotation.

Without the other effects the deformed potential generates a set of energy levels which are filled by neutrons and protons up to some 'Fermi energy'. In a heavy nucleus many of these levels correspond to states of high angular momentum, though for a nucleus with an even number of neutrons and protons the ground-state spin is invariably zero. This is a result of the attractive residual interactions which cause the nucleons with energies near the Fermi surface to couple into pairs which, classically speaking, move around the nucleus in the same orbits but in opposite directions, and always scatter into other unoccupied 'time-reversed' states (see Fig. 2a). This correlation reduces the total energy since the paired nucleons now experience their mutual attraction on every circuit. While the nuclear shape and average field rotate, the individual nucleons move around in the interior with velocities generally considerably higher than the collective velocities due to the rotation. The fact that many of the paired nucleons tend to move in the opposite direction to the overall rotation causes the nucleus to behave in a superfluid manner and this reduces $\mathcal{J}$ at low ω .

The nucleons, however, also experience a Coriolis interaction. This is frequently referred to as an apparent force since it describes deviations in the motion of an object when viewed from a rotating reference frame.

Consider, for example, a satellite orbiting the Earth. Since the Earth is roughly spherical, the motion of the satellite is independent of the Earth's orientation, though for non-equatorial orbits the simple planar motion looks a little peculiar when projected on to the Earth's rotating surface. However, the Coriolis force also has very real dynamical consequences. For example it causes the rotation of air masses which we know as cyclones or anticyclones. The reason for the difference between these real and apparent effects is that as far as the air mass is concerned, the Earth does not possess spherical symmetry since there are certain 'privileged' points of high and low

pressure. These rotate with the Earth and since the air flows towards a moving low-pressure area, a stable circulating wind pattern may result.

Rotational nuclei are also non-spherical and the Coriolis force again affects the way the motion of individual nucleons couples to the rotations. Ultimately the inertial forces make the 'aligned' configuration, depicted classically in Fig. 2b, energetically more favourable and the intrinsic configuration flips over at ω_c — with a consequent increase in both the spin and $\mathcal{J}$. (The phenomenon is formally very similar to the breaking of electron pairs in a superconductor by the application of an external magnetic field.)

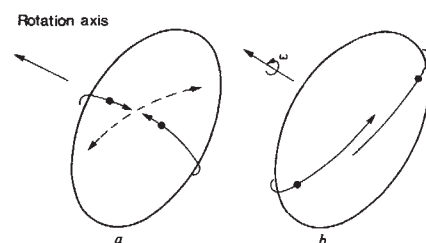


Fig. 2 Classical representation of the nuclear configuration below (a) and above (b) the crossing frequency.

In quantum-mechanical terms the nuclear state in the region of ω_c will be an admixture of both these configurations. The admixture is essentially determined by how effective the Coriolis force is in deflecting nucleons from orbits occupied in the paired phase to those occupied in the unpaired phase. Ironically, a spectacular backbend is obtained if the Coriolis mixing is weak because the cross-over will then only happen suddenly as the pure aligned configuration attains the lower energy. The strength of the Coriolis mixing has been seen in the rare earths to be an oscillating function of neutron number, with a consequent alternation between upbending and backbending. This is essentially due to the different nature of the states occurring near the Fermi surface in nuclei with more or less neutrons (R. Bengtsson *et al.* *Phys. Lett.* **73B**, 259; 1978).

The observation of backbending in ^{244}Pu (where alignment is apparently due to protons) confirms a similar behaviour in the actinides. More importantly perhaps, the existence of a sharp backbend allows a precise determination of the crossing frequency and of the associated increases in spin and moment of inertia, enabling a quantitative comparison with current theories to be made on these heavy systems. As yet, even the data of Spreng *et al.* do not span the entire backbending region. But there is no doubt that they will stimulate further spectroscopy in this mass region. □

Neil Rowley is in the Theory and Computational Science Division at the Science and Engineering Research Council Daresbury Laboratory, Daresbury, Warrington WA4 4AD.

Immunology

Induction of tolerance to self in T lymphocytes

from Ron Schwartz

THE fundamental problem confronting the immune system is to discriminate between the 'self' molecules that constitute the organism and the 'nonself' molecules on foreign organisms, tissues or substances that challenge it. One of the central dogmas of immunology is that the ability to discriminate between self and nonself is mainly acquired during the development of the lymphoid system and results in tolerance — the lack of responsiveness — to self molecules. Two major mechanisms for the induction of self tolerance have been suggested: clonal deletion (anergy), which postulates that clones of lymphocytes specific for self antigens are selectively destroyed or functionally inactivated by exposure to the antigens either in the absence of a second critical differentiation signal or simply early in the cells' development¹; and an active regulatory mechanism, in which negative feedback signals from other cells — the so-called suppressor T cells — in the immune system prevent self-recognizing clones from responding to self antigens².

For the antibody-producing B lymphocyte both mechanisms are probably operative to different degrees depending on the particular self-antigen being recognized³ (although it has been argued by one group that tolerance induction does not occur at all at the level of the B lymphocyte⁴). For the T lymphocyte, very little is known about the mechanism of tolerance induction although a number of recent studies, including two reported on pages 738 and 741 of this issue of *Nature*, have made some progress — if only to refute a major prediction of the only clearly espoused theory of the induction of T-cell tolerance⁵.

The Cohn-Epstein theory represents part of an extension to T lymphocytes of the Bretcher-Cohn two-signal model for B-lymphocyte activation and is based on the widely held belief that, in order to be activated, a T cell must recognize both foreign antigen and a gene product of the major histocompatibility complex (MHC) found on the surface of the antigen-presenting cell⁶. (This has been termed MHC restriction since the activation of any T cell by an antigen-presenting cell is restricted to cells that share a given allelic form of the MHC-encoded molecule.) In the Cohn-Epstein model, each T cell has two combining sites, one specific for antigen, the other for the MHC-encoded molecule. If both sites are occupied, the T cell is activated; if only the antigen-specific site is occupied, the T cell receives an 'off' signal. Thus, in this model, the induction

of tolerance is explained by an encounter between a self antigen and a T cell in the absence of MHC-encoded molecules early in development. A clear prediction of the Cohn-Epstein model is that tolerance induction should not be MHC-restricted, that is, an animal should not respond to a self antigen no matter what allelic form of the MHC-encoded molecule is expressed on the antigen-presenting cell. This prediction has now been tested in a number of ways and shown to be false.

The initial observations were made with T-cell clones and lines specific for a given antigen and the first example was published by Dos Reis and Shevach⁷. They isolated from strain 13 guinea pigs proliferative (presumably helper) T cells specific for beef insulin in association with a foreign (strain 2) MHC-encoded molecule — a class II histocompatibility antigen. This was done by first depleting the T-cell population of a strain 13 guinea pig of cells specific for strain 2 lymphoid cells alone and then stimulating the remaining strain 13 T cells with a mixture of strain 2 antigen-presenting cells and beef insulin. That generated T cells with a specificity for insulin and the allogeneic strain 2 class II molecules. Most importantly, Dos Reis and Shevach went on to show that the antigenic site on the insulin molecule recognized by such T cells was localized to the B chain, between residues 5 and 16. Much to their surprise, since strain 13 animals do not respond when immunized with guinea pig insulin, the same determinant was found on guinea pig insulin, implying that the animals were tolerant to guinea pig insulin only in association with their own class II molecules. This suggested that the induction of tolerance was MHC-restricted.

These findings have now been extended to mouse cytotoxic T lymphocytes and class I antigens of the murine MHC (*H-2*) by Rammensee and Bevan⁸. Using a protocol similar to that of Dos Reis and Shevach, they depleted a spleen-cell preparation from BALB mice, which are *H-2^d*, of cells specific for *H-2^k* molecules (using cells from B10.BR mice) and then developed cytotoxic T-cell lines specific for a minor surface antigen of their own in association with *H-2^k* molecules. Such lines, isolated on two separate occasions, were shown to be specific for the combination of BALB minor and *H-2^k*. Therefore, since BALB mice never make a cytotoxic response against themselves (BALB minor plus *H-2^d*), they must be tolerant to BALB minor only in the context of their own class I molecules.

One problem with the *in vitro* studies of Dos Reis and Shevach and of Rammensee and Bevan is that they fail to eliminate the possibility that somatic mutation has occurred in the structural genes encoding the T-cell receptor during the cell-cloning process. This objection is not applicable, however, to a series of elegant *in vivo* experiments, involving radiation-induced bone-marrow chimaeras, carried out by Groves and Singer⁹.

They created chimaeras by reconstituting the bone marrow of irradiated mice that were the first-generation offspring of crosses between B10 (*H-2^b*) and B10.BR (*H-2^k*) animals with marrow from C3H.SW (*H-2^b*) mice. Although the donor is *H-2^b*, when its cells develop in the B10 × B10.BR mice their restriction specificity is expanded so that they can also recognize foreign antigens in association with *H-2^k* molecules. The question which Groves and Singer asked was whether the chimaeric T cells would recognize C3H minor antigens, which are expressed on the donor but not the recipient cells, in association with *H-2^k*. They should not if the induction of tolerance is unrestricted. But they did. In quantitative terms, the cytotoxicity against C3H minor antigens seemed less than that generated by normal (B10 × B10.BR)F₁ mice. Nonetheless, it was clear that at least a portion of the T-cell repertoire specific for C3H minor antigens was tolerized in a MHC-restricted manner.

That conclusion is now confirmed by Matzinger *et al.*¹⁰, using another *in vivo* model, thymus-grafted chimaeric mice. (B10 × B10.D2)F₁ mice were thymectomized, lethally irradiated, reconstituted with F₁ bone marrow and then grafted with a thymus bearing BALB minor antigens and *H-2^b* molecules. Matzinger *et al.* found that cytotoxic T cells from the animals were tolerant to BALB minor antigens in association with *H-2^b* molecules but not to BALB minor antigens in association with *H-2^d*, again demonstrating MHC restriction of the induction of T-cell tolerance.

Finally, Lowy *et al.*¹¹ have been led to the same conclusion using a completely different *in vivo* model in which normal adult mice are made unresponsive to the hapten azobenzene arsonate (ABA) by injecting them intravenously with ABA-coupled syngeneic spleen cells. (This manoeuvre prevents the induction of a delayed-type hypersensitivity response when these animals are subsequently immunized subcutaneously with the ABA-coupled cells.) An analysis of the nonresponsive state revealed two components, one of which could be eliminated by pretreatment of the animals with cyclophosphamide or pretreatment of the hapten-coupled spleen cells with anti-I-J antibody and complement, the other of which was resistant to these manipulations and not transferable to naive animals. The

interesting result was that for induction of the latter type, the ABA-coupled spleen cells were required to share certain class II molecules with the recipient — that is, the induction event was MHC-restricted. Thus, even in a non-ontogenic model, for at least one limb of the tolerance-inducing process, the cells involved must recognize antigen in association with MHC-encoded gene products.

This suggests that the distinction of self from nonself is not mediated by any event in ontogeny involving the expression or recognition of only a part of the T-cell receptor, for example only one of its two polypeptide chains¹². If the two-chain molecule forms a single combining site, similar to that of heavy and light chains for immunoglobulin, one can postulate that the ability to distinguish between self and nonself might in some cases result solely from the interactions between the antigen and the histocompatibility molecule. A recent example which seems to fulfill this prediction is the T-cell proliferative response to pigeon cytochrome *c* in the B10.A mouse^{13,14}.

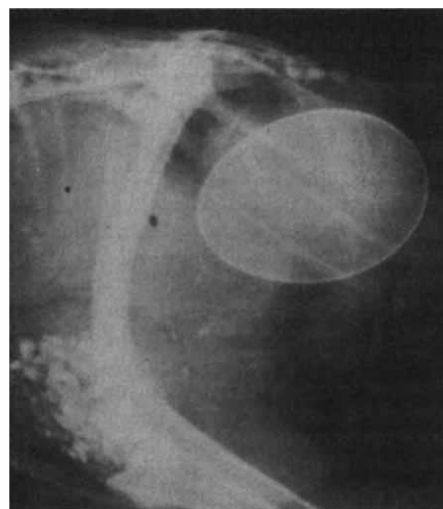
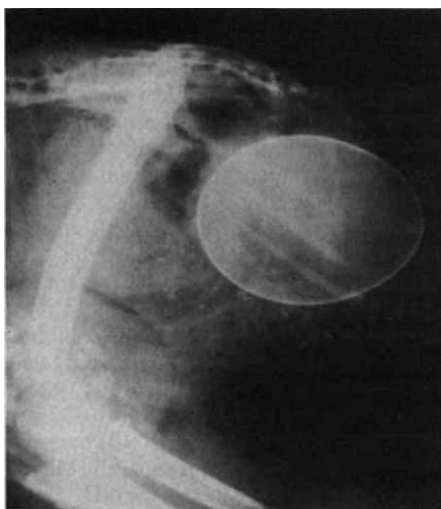
Structural studies using synthetic peptide analogues have suggested that a lysine at position 99 of cytochrome *c* is a contact residue for the T-cell receptor whereas a C-terminal lysine at position 103 or 104 is important for interaction with the class II histocompatibility molecule. A comparison of pigeon and mouse cytochrome *c* sequences reveals that both possess a lysine at 99 but that whereas the pigeon has Ala 103, Lys 104, the mouse has Asn 103, Glu 104. Thus mouse cytochrome *c* fails to stimulate T-cell clones from B10.A mice specific for pigeon cytochrome *c* solely on the basis of sequence differences that affect its interaction with the class II molecule. T-cell clones specific for mouse cytochrome *c* in association with B10.A class II molecules are presumably deleted or suppressed by the self-tolerance mechanism.

Overall, these experiments involving *in vitro* and *in vivo* experimental models of helper, cytotoxic and delayed-type hypersensitivity T-cell function have led to one and the same conclusion: that the induction of T-cell tolerance is MHC-restricted. This directly refutes a major prediction of the Cohn-Epstein model of T-cell activation and suggests that

independent recognition of antigen without histocompatibility molecules does not occur even during ontogeny. These results do not, however, eliminate the application to T cells of the Bretcher-Cohn two-signal model of discrimination between self and nonself. Thus it is possible that engagement of the T-cell receptor by both antigen and histocompatibility molecule delivers signal one which, by itself, induces tolerance, but if

coupled with a second signal (for example interleukin 1), transduced through a separate receptor on the T-cell membrane, induces activation. Such models are currently being tested in several laboratories using anti-receptor antibodies and T-cell clones^{15,16}. □

Ron Schwartz is in the Laboratory of Immunology, Building 10, Room 11N311, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20205.



AVIAN eggs are laid blunt end first. For many years it was taken for granted that birds form eggs that way, perhaps first creating soft spheroids and then gradually deforming them into ovoids by the peristaltic movement of their uterine muscles. Although several physiologists, starting with Fabricius of Padua in the 17th century, realized that the egg rotates *in utero*, thereby confuting the peristaltic theory, confusion abounded until J.R.G. Bradfield, in the Department of Zoology at Cambridge, asked J.A.F. Fozzard, an experienced radiographer at the School of Anatomy, to X-ray hens at successive stages of their incubation cycle. In April 1946, Fozzard photographed a Rhode Island Leghorn with a Coolidge X-ray tube at 60 kV placed 1 m away and obtained the radiographs shown here with exposures of 0.05 s.

The faint outline of an egg in production becomes discernable about a quarter of the way through the 26-hour cycle of production and by half way though the cycle it is unmistakably oval. Until 23.5 hours (above left) the sharp end points towards the cloaca but by 25.5 hours (above right) rotation of the egg has directed the blunt end towards the cloacal exit. In between, the brooding hen rises to her feet to enable the egg to rotate and drop to a deeper position.

So the ingenious explanation offered for the egg's form by D'Arcy Thompson in his famous work *On Growth and Form* — that muscular contractions propelling an egg down the oviduct account for its shape — was radiographically refuted.

Why should birds behave in this extra-

ordinary fashion? Obviously there must be some natural advantage to the blunt-laid egg, and in an article in the *British Medical Journal*, 1st vol, p.585 (1948), entitled "De formatioine ovi et pulli" in imitation of Fabricius and William Croone, H.A. Harris advanced one theory. The peristaltic uterine contractions of a hen are surprisingly forceful, he argued, and the blunt-end-foremost arrangement is favoured to avoid rupture of either egg or uterus. However, Bradfield found the alleged peristalsis to be without foundation (*J. exp. Biol.* 28, 125, 1951). Instead, the shell gland holding the egg is extruded through the cloaca where the egg is deposited before the gland retracts. Is there some other reason why hen eggs are ovoid? If the ovoid form has such exceptional tensile strength, why are some birds' eggs spherical? Are all birds equipped with an internal rotation system to guarantee their eggs a blunt exit? And why is the egg not formed the 'right' way round in the first place?

In the decades since Bradfield's study, various papers have referred to his analysis of the rate of calcium deposition in egg-shell formation, but none has accounted satisfactorily for the purpose and mechanism of rotation. Could this long-standing problem be met with a seasonal solution from an Easter reader of the last of this series of historic scientific photographs?

Jon Darius

Taken from *Beyond Vision*, an exhibition just opened at the Science Museum, London and a book of the same title just published by Oxford University Press. Reproduced by kind permission of J.A. Fairfax Fozzard.

- Weigle, W. O. *Adv. Immun.* 16, 61 (1973).
- Nossal, G. J. V. *A. Rev. Immun.* 1, 33 (1983).
- Waters, C. A. & Diener, E. *Eur. J. Immun.* 13, 928 (1983).
- Fernandez, C. *et al. Immun. Rev.* 43, 1 (1978).
- Cohn, M. & Epstein, R. *Cell. Immun.* 39, 125 (1978).
- Bretcher, P. & Cohn, M. *Science* 169, 1042 (1970).
- Dos Reis, G. A. & Shevach, E. M. *J. exp. Med.* 157, 1287 (1983).
- Rammensee, H.-G. & Bevan, M. J. *Nature* 308, 741 (1984).
- Groves, E. S. & Singer, A. *J. exp. Med.* 158, 1483 (1983).
- Matzinger, P., Zamojska, R. & Waldmann, H. *Nature* 308, 738 (1984).
- Lowy, A., Debrin, J. A., Monroe, J. G., Granstein, R. D. & Greene, M. I. *Nature* 308, 373 (1984).
- Hedrick, S. M. *et al. Nature* 308, 153 (1984).
- Hansburg, D. H. *et al. J. Immun.* 131, 319 (1983).
- Hansburg, D. H. *et al. J. exp. Med.* 158, 25 (1983).
- Meuer, S. C. *et al. J. exp. Med.* 158, 988 (1983).
- Kaye, J. *et al. J. exp. Med.* 158, 836 (1983).

Planetary science

Ices in the Solar System

from I. P. Wright

THE growth of interest in Solar System ices has been generated, in part, by visually spectacular data from the Voyager encounters with the rings and satellites of Jupiter and Saturn. At the same time, substantial advances have been made with theoretical modelling and observational measurements of dirty ices, long thought to be the primary constituents of cometary nuclei, and there is intensive laboratory study of the physical properties of ice. These and other topics were the subject of a meeting recently held in Nice*.

Water ice is thought to be prevalent in certain satellites of Jupiter and Saturn as well as in comets. E. Whalley (National Research Council, Ottawa) reviewed those properties of water ice (or the 'fundamentals of planetary glaciology' as he preferred to call it) that are pertinent to the study of large icy bodies. The phase diagram of water over temperatures between 0 and -120°C and pressures up to 25 kbar shows nine different polymorphs of ice (called ice I_h , ice II, ice III and so on), some of which are ordered structures, others disordered. A novel ordering transformation identified in ice I_h at low temperatures could possibly occur in an icy Solar System body, catalysed by such substances as ammonia. If so, since the transformation liberates heat, it would have a number of implications for the physical states of planetary bodies.

The water ice deep in the interiors of the satellites of Jupiter and Saturn may be the high-pressure forms, ice VI and VII. J. Poirier (Institut de Physique du Globe de Paris) described the use of a sapphire-anvil squeezer cell to study the laboratory deformation of these forms. By measuring the displacement of markers within the ice crystals it was possible to compute the creep rates of ice as a function of shear stress, temperature and pressure; preliminary measurements of ice VII show it to be extremely hard. Experimental deformation of ice I_h demonstrates that it can exhibit both brittle and ductile behaviour (S. Kirby, US Geological Survey, Menlo Park; W. Durham and H. Heard, Lawrence Livermore National Laboratory). Its brittle character is similar to that of rocks; measurements of ductile strengths, however, indicate that the ice I_h layer on icy Solar System bodies is much weaker than generally thought.

Of relevance to studies of ice accumulation was the report by E. Mayer and R. Pletzer (Universitat Innsbruck, Austria) that the properties of amorphous ice are dependent upon the method of production. In one of their methods, water

vapour is leaked through a nozzle into a high-vacuum chamber and deposited as amorphous ice on a cryoplate. The ice, formed in supersonic flow conditions, is found to have a very rough surface; upon warming it shows a number of exothermic peaks. In their alternative method, the water flow is broken up by the introduction of a baffle, whereupon a glass-like ice with a very small surface area is formed. This ice produces only one exothermic peak when warmed. The interesting question that remains is what form of ice might be produced when a solid target travels at high velocity through water vapour, the probable situation in certain Solar System environments.

M. Nicol and co-workers (University of California, Los Angeles) have investigated ice formation in an ammonia-water system at low ammonia concentrations (0–30 per cent) and high pressures (up to 40 kbar), using a diamond-anvil squeezer cell. Microscopy showed that whereas the $\text{NH}_3 \cdot 2\text{H}_2\text{O}$ phase is characterized by strong anisotropy, smaller refractive index than ice VII and the tendency to form cracks, the $\text{NH}_3 \cdot \text{H}_2\text{O}$ phase, which forms at higher pressures, has a refractive index close to that of ice VII and does not form cracks. The outcome of the work has been the production of a preliminary phase diagram for the ammonia-water system. The importance of this system was outlined by S. Miller (University of California, San Diego) who believes that hydrates of ammonia are likely to occur on Uranus and Neptune as well as on the satellites of Jupiter and Saturn.

Accompanying the ammonia hydrates on icy Solar System bodies might be methane hydrate, a substance known to occur on Earth in certain deep-sea sedi-

ments. Furthermore, by analogy with the air hydrates found in the terrestrial Antarctic ice cap, it is thought that CO_2 hydrates should be present in the polar regions of Mars (and may also exist in comets). Clathrate hydrates are thus thought to be important and widespread constituents of Solar System ices and now appear in theoretical models of the physical and chemical states of the saturnian satellite Titan and the neptunian satellite Triton (J. Lunine, Caltech). T. Owen (Stony Brook, State University of New York) demonstrated that the atmospheres of icy bodies are composed of gases which probably result from the dissociation of clathrates. Unfortunately clathrate hydrates have observational characteristics which are indistinguishable from water ice and so, without an actual excursion to one of the icy Solar System bodies, it will not be possible to verify their existence anywhere other than on Earth.

An important constituent of bodies such as Pluto, Titan and Triton is methane ice. Indeed, on Triton there is evidence of islands of solid methane floating on a sea of liquid nitrogen (D. Cruikshank, University of Hawaii). By way of complimenting the observational work, N. Trappeniers (University of Amsterdam) reported that pressures of up to 60 kbar, attained in a diamond-anvil cell, produce evidence of a new phase of methane (phase IV).

All told, the meeting demonstrated that definite progress is being made towards the ultimate goal — a full understanding of the evolution of ices in the Solar System and the construction of consistent models which relate the nature and characteristics of the ices to the materials from which they are formed. It was easy to sense the impatient expectation of many participants for the unique measurements that should come from fly-by missions to Halley's comet when it makes its closest approach to Earth in 1985–6. □

I. P. Wright is in the Department of Earth Sciences, The Open University, Milton Keynes MK7 6AA.



100 years ago

AMONG the superabundant "Universities" of the United States Harvard is unquestionably taking its place as a national institution on a par with British establishments which hold a similar designation. The last quarterly *Bulletin* of its proceedings is before us, which has to acknowledge during that short time nine legacies or donations in money, varying from 200 to 100,000 dollars, and amounting to 168,000 dollars. One of these is 10,000 dollars subscribed for the purchase of meteorites, and another is 2000 dollars from the Massachusetts Society for Promoting Agriculture, to assist in the establishment of a veterinary hospital, to which institution also a collection of pathological models is presented.

THE last number of the *Transactions of the Seismological Society of Japan* (Yokohama, 1884) contains various papers on seismology. The first is by Prof. Milne, on earth pulsations; the next is by Mr. Alexander, on the interpretation of a diagram described by a particular form of earthquake instrument. The object of the writer is to calculate not only the maximum velocity, but also the maximum rate at which the velocity changes, "which is a measure of the effect which an earthquake exerts in overturning and fracturing bodies placed on the earth's surface." Prof. Ewing describes the construction of a pendulum which shall be without a tendency to swing when the point from which it is suspended suffers displacement. Mr. Gergens gives a note on ripple-like marks found on the surface of an iron casting supposed to have been shaken while solidifying, which marks are picturesquely described as "a note in a congealed earthquake." The remainder of the volume contains suggestions for new types of seismographs. From *Nature* 29, 583, 17 April 1884.

Ices in the Solar System, a NATO-sponsored Advanced Research Workshop, was held in Nice from 16 to 19 January.

The role of protein kinase C in cell surface signal transduction and tumour promotion

Yasutomi Nishizuka

Department of Biochemistry, Kobe University School of Medicine, Kobe 650, Japan, and Department of Cell Biology, National Institute for Basic Biology, Okazaki 444, Japan

Protein kinase C has a crucial role in signal transduction for a variety of biologically active substances which activate cellular functions and proliferation. When cells are stimulated, protein kinase C is transiently activated by diacylglycerol which is produced in the membrane during the signal-induced turnover of inositol phospholipids. Tumour-promoting phorbol esters, when intercalated into the cell membrane, may substitute for diacylglycerol and permanently activate protein kinase C. The enzyme probably serves as a receptor for the tumour promoters. Further exploration of the roles of this enzyme may provide clues for understanding the mechanism of cell growth and differentiation.

THE biochemical basis of the transduction of extracellular signals into intracellular events and of cellular proliferation has long been a subject of great interest. Recently, much attention has been paid to the role in such events of the turnover of inositol phospholipids in the membrane that is commonly observed immediately after the interaction of a wide variety of biologically active substances with their specific cell surface receptors. This turnover of inositol phospholipids is associated with an increase in intracellular calcium ion, which appears to mediate many of the subsequent physiological responses.

When the enzyme, protein kinase C, was first found, in 1977, as a proteolytically activated protein kinase in many tissues, it had no obvious role in signal transduction¹. Later, it was shown to be a Ca^{2+} -activated, phospholipid-dependent enzyme², and was firmly linked to signal transduction by the demonstration that diacylglycerol, one of the earliest products of signal-induced inositol phospholipid breakdown, greatly increased the affinity of protein kinase C for Ca^{2+} , thereby activating it³⁻⁶. The enzyme apparently possesses multifunctional catalytic activity.

The purpose of this review is to summarize the evidence relating to the activation of protein kinase C and to the importance of that activation for subsequent cellular responses. It will be argued that the activation of cellular responses by this route is separate from and synergistic to those activated via an increase in intracellular Ca^{2+} as outlined in Fig. 1. Recent evidence that protein kinase C is a prime target for the actions of phorbol ester and perhaps some other tumour promoters will also be reviewed. Several other aspects of protein kinase C have been reviewed elsewhere⁷⁻¹¹.

Biochemical activation and properties

Protein kinase C is a Ca^{2+} - and phospholipid-dependent enzyme that is activated by diacylglycerol^{3,4}. Diacylglycerol is normally almost absent from membranes, but is transiently produced from inositol phospholipids in response to extracellular signals. Kinetic analysis indicates that a small amount of diacylglycerol dramatically increases the apparent affinity of protein kinase C for Ca^{2+} , fully activating the enzyme without any change in Ca^{2+} levels^{4,12}. Triacylglycerol, monoacylglycerol and free fatty acid are totally inactive. The active diacylglycerols appear to contain at least one unsaturated fatty acid irrespective of the chain length of the other fatty acyl moiety, but the precise structural requirements for activity have not been extensively investigated. For reasons that will become apparent later, it is worth noting that inositol phospholipids in mammalian tissues frequently contain arachidonic acid at position 2. In the activation of protein kinase C, phosphatidylserine is indispensable,

and other phospholipids show positive or negative cooperativity with phosphatidylserine¹². For example, when phosphatidylethanolamine is present as an additional lipid component, the enzyme is fully activated by diacylglycerol at the 10^{-7} M range of Ca^{2+} , whereas phosphatidylcholine and sphingomyelin are inhibitory. Therefore, the asymmetric distribution of various phospholipids in the lipid bilayer may play a part in enzyme activation. The enzyme is independent of calmodulin, but is profoundly inhibited by a large number of phospholipid-interacting drugs including chlorpromazine, dibucaine and trifluoperazine, which are known calmodulin inhibitors^{13,14}.

Protein kinase C is widely distributed in tissues and organs of mammals and other organisms^{1,5,15,16}. In many tissues its activity far exceeds that of cyclic AMP-dependent protein kinase (protein kinase A) when assayed with calf thymus H1 histone as a common substrate. It consists of a single polypeptide chain (molecular weight 77,000) that appears to be composed of two functionally different domains which can be separated with Ca^{2+} -dependent thiol proteases^{16,17}. One is a hydrophobic domain that may bind to membranes; the other is a hydrophilic domain that carries the catalytically active centre. The latter fragment is fully active without Ca^{2+} , phospholipid and diacylglycerol. (A class of Ca^{2+} -dependent thiol proteases which are active at the 1–10 μM range of Ca^{2+} preferentially cleave the activated form of protein kinase C¹⁷; the physiological significance of this irreversible reaction is unknown.)

When tested *in vitro*, protein kinase C has a broad substrate specificity, phosphorylating seryl and threonyl residues, but not tyrosyl residues, of many endogenous proteins in most tissues^{5,9-11}. Protein kinases C and A share many phosphate acceptor proteins *in vitro* but with very different relative rates of phosphorylation of different amino acyl residues^{5,10}. No doubt the primary sequence in the vicinity of the phosphorylation sites is an important factor in determining substrate recognition by different protein kinases, as is the topographical arrangement or subcellular localization of the enzymes and their substrates (for reviews see refs 5,18,19). The nature of the physiological substrates of protein kinase C in most tissues remains largely unexplored.

Physiological activation

A wide variety of extracellular signals which activate cellular functions and proliferation through interaction with membrane receptors have repeatedly been shown to provoke the breakdown of inositol phospholipids in their target tissues (for reviews see refs 20–22). Originally phosphatidylinositol was considered to be the target²³, but recently it has been proposed that, in

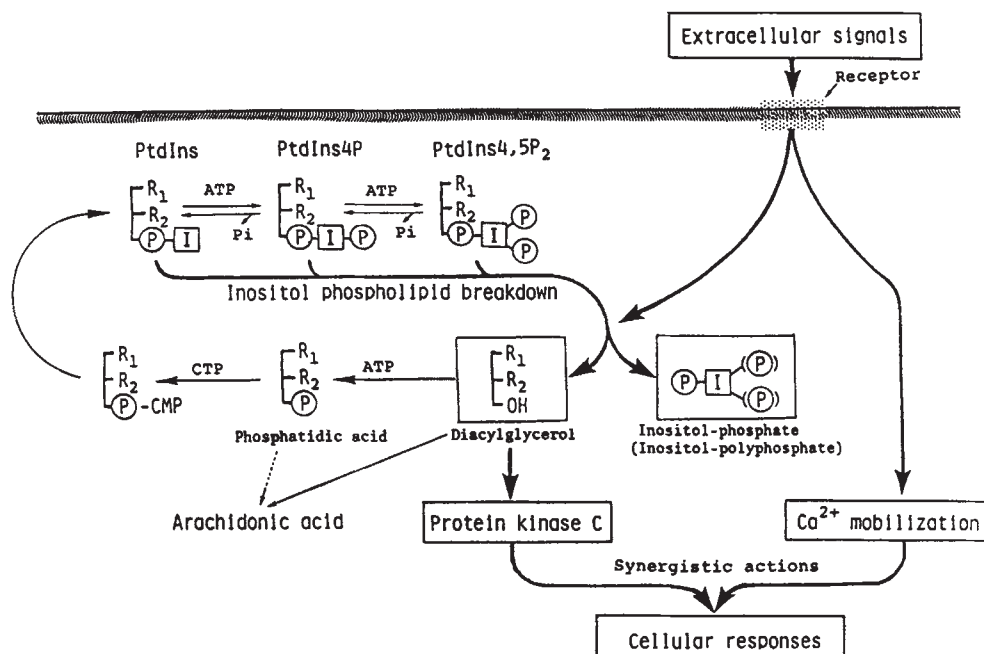


Fig. 1 Inositol phospholipid turnover and signal transduction. PtdIns, phosphatidylinositol; PtdIns4P, phosphatidylinositol-4-phosphate; PtdIns4,5P₂, phosphatidylinositol-4,5-bisphosphate; R₁ and R₂, fatty acyl groups; I, inositol; and P, phosphoryl group.

some mammalian tissues (for a review see ref. 24) and insect salivary gland²⁵, the immediate target is phosphatidylinositol biphosphate even though this phospholipid is a very minor component in membranes (Fig. 1). The evidence so far presented is plausible, but the precise physiological picture and tissue variations may be established by further investigations. In any case, the primary products of the reaction are diacylglycerol and inositol phosphate or inositol polyphosphate²⁶. Evidence that this signal-induced reaction is directly linked to activation of protein kinase C has primarily come from experiments with platelets for which there are many useful agonists and antagonists^{6,27,28}.

In platelets stimulated by thrombin²⁷⁻³⁰, collagen²⁸ or platelet-activating factor³¹, two endogenous proteins with approximate molecular weights of 40,000 (40K protein) and 20,000 (20K protein) are rapidly and heavily phosphorylated in association with release of various constituents of platelet granules such as serotonin from dense bodies and acid hydrolases from lysosomes. The 20K protein is myosin light chain: calmodulin-dependent phosphorylation of this protein absolutely requires mobilization of Ca²⁺ (ref. 32). On the other hand, the 40K protein isolated from human platelets is a substrate for protein kinase C²⁸. More recently, Imaoka *et al.*³³ have confirmed this reaction using a highly purified preparation of the 40K protein, which they find has a 47K component with some microheterogeneity on isoelectrofocusing electrophoresis. The function of the 40K protein remains unknown.

It has been repeatedly shown that, when stimulated, platelets rapidly produce diacylglycerol containing arachidonic acid, and that this reaction is accompanied by the concomitant disappearance of inositol phospholipids^{27,28,31,34,35}. It is important to note that diacylglycerol is only transiently produced in membranes, presumably due both to its conversion back to inositol phospholipids and to its further degradation to arachidonic acid for thromboxane and prostaglandin synthesis (Fig. 1). The transient appearance of diacylglycerol in membranes is always associated with activation of protein kinase C as judged by 40K protein phosphorylation^{27,28,31}.

In the *in vitro* enzymatic reaction, various synthetic diacylglycerols are equally capable of activating protein kinase C. Although those of the synthetic diacylglycerols that contain two long fatty acyl moieties such as diolein are poor inducers of 40K protein phosphorylation in intact cell systems, presumably due to their inability to be intercalated into the membrane, others are able to activate protein kinase C in intact cells^{6,36,37}. For instance, when 1-oleoyl-2-acetyl-glycerol (shown in Fig. 2) is

suspended in 1% dimethyl sulphoxide and added to intact platelets directly, the 40K protein is rapidly and fully phosphorylated but without any concomitant breakdown of inositol phospholipids, formation of diacylglycerol, arachidonic acid release or thromboxane synthesis^{36,37}. There is no sign of damage to the cell membranes. This diacylglycerol does not appear to interact with the receptor for platelet-activating factor, since the latter factor definitely induces rapid breakdown of inositol phospholipids³¹. Instead, the synthetic diacylglycerol is shown to be metabolized rapidly *in situ* to its corresponding phosphatidic acid, 1-oleoyl-2-acetyl-3-phosphorylglycerol, presumably by the action of diacylglycerol kinase³⁷. Fingerprint analysis of the tryptic phosphopeptides of the 40K protein that is phosphorylated in platelets treated with the synthetic diacylglycerol confirms that the phosphorylation is catalysed by protein kinase C. On the basis of these observations, together with evidence obtained with inhibitors such as chlorpromazine and dibucaine⁷, it may be concluded that protein kinase C is indeed activated by extracellular signals in physiological processes, and that diacylglycerol, transiently produced from inositol phospholipids in response to the extracellular signals reaching their target cell membranes, serves as the direct activator of this protein kinase.

Activation by tumour promoters

Several phorbol esters such as 12-O-tetradecanoylphorbol-13-acetate (TPA) are well known as potent tumour promoters, and several kinetic studies with various cell types suggest that their primary site of action is the cell surface membrane (for review see refs 38-40). Recent studies in our laboratory have provided evidence that protein kinase C is a target for phorbol esters, as the tumour promoters directly activate this enzyme both *in vitro* and *in vivo*, and there is an approximate correlation between the ability of individual phorbol esters to promote tumours and to activate the enzyme^{41,42}. Kinetic analysis indicates that TPA, which has a diacylglycerol-like structure (Fig. 2), is able to substitute for diacylglycerol at extremely low concentrations. Like diacylglycerol, TPA dramatically increases the affinity of the enzyme for Ca²⁺ to the 10⁻⁷ M range, resulting in its full activation without detectable cellular mobilization of Ca²⁺ (ref. 42). The binding of ³H-labelled phorbol-12,13-dibutyrate (PDBu), which is a potent tumour promoter but less hydrophobic than TPA, to purified protein kinase C has an absolute requirement for Ca²⁺ and phospholipid⁴³. ³H-PDBu will bind neither to protein kinase C nor to phospholipid *per se* irrespective of the presence or absence of Ca²⁺; only if all four components are present simultaneously do binding and enzyme

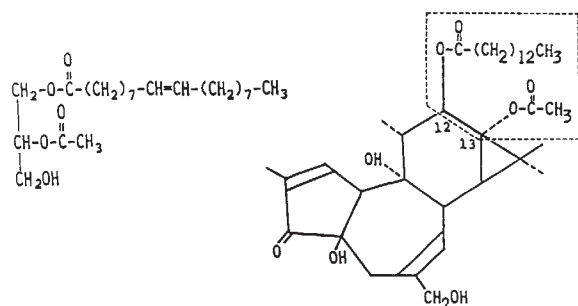


Fig. 2 Structures of synthetic diacylglycerol (1-oleoyl-2-acetyl-glycerol; left) and tumour-promoting phorbol ester (1-O-tetradecanoylphorbol-13-acetate, TPA; right). TPA contains a diacylglycerol-like structure in its molecule, as shown in the dotted square.

activation occur. The effects of varying the concentration of PDBu on the activation of protein kinase C and on its binding to the enzyme are shown in Fig. 3. The apparent dissociation binding constant (K_d) of the tumour promoter is estimated to be 8 nM, which is the same as the activation constant (K_a) for protein kinase C calculated by Lineweaver-Burk double reciprocal plots. These values, obtained with homogeneous protein kinase C and chromatographically pure phosphatidylserine, are remarkably similar to the K_d values described for the specific tumour promoter-binding site in brain particulate fractions (5.6 nM)⁴⁴, and on intact cell membranes of rat embryo fibroblasts (8 nM)⁴⁵ and mouse epidermal cells (10 nM)⁴⁶. Recent reports from several laboratories⁴⁷⁻⁵⁰ have also confirmed that both Ca^{2+} and phospholipid are needed for the binding of tumour promoter to a specific binding site that is presumably protein kinase C.

Scatchard analysis indicates that roughly one molecule of PDBu binds to one molecule of protein kinase C in the presence of a physiological concentration of Ca^{2+} and an apparent excess of phospholipid (Fig. 3)⁴³, but this stoichiometry does not necessarily indicate that the tumour promoter binds directly to the enzyme. Using a phorbol ester photoaffinity probe Delclos *et al.*⁵¹ have obtained evidence that the phorbol ester interacts primarily with phospholipid. The tumour promoter activates the enzyme presumably by some modification of its phospholipid microenvironment. The phospholipid specificity for PDBu binding is similar to that for the activation of protein kinase C, with phosphatidylserine being most effective^{43,47-50}.

The amount of phospholipid needed for binding in *in vitro* assays, for example 100 $\mu\text{g ml}^{-1}$ phosphatidylserine, usually far exceeds the amount of protein kinase C and of tumour promoter⁴³. Although it is difficult to translate this concentration into physiological terms, since phospholipid is water-insoluble, it is nevertheless attractive to suggest that, in intact cells where phospholipid and Ca^{2+} are not limited, each molecule of tumour-promoting phorbol ester which intercalates into the membrane phospholipid bilayer interacts with one molecule of protein kinase C which moves towards it to produce the quaternary complex by which the enzyme is activated; a new concept of the 'mobile receptor'. Consistent with this proposal is the fact that, in resting conditions in most tissues, protein kinase C is largely present in a soluble inactive form but when, for example, TPA or PDBu is added to intact cells, protein kinase C is recovered in a form tightly associated with cell membranes on disruption of the cells (see also ref. 52).

The results obtained with tumour promoters raise the possibility that, in physiological processes, one molecule of diacylglycerol, produced from inositol phospholipids in the membrane, activates one molecule of protein kinase C. If so, it may be that even though phosphatidylinositol bisphosphate is a very minor membrane component, its breakdown provides sufficient diacylglycerol to activate all the protein kinase C in the cell. However, it is difficult at present to imagine the mechanism by

which protein kinase C senses such a tiny change in the constitution of cell membranes.

Further evidence supporting the view that protein kinase C acts as the receptor for tumour-promoting phorbol ester, comes from the similar distribution patterns among tissues of the enzyme and phorbol ester binding sites^{43,47,53}. It is also noteworthy that Nield *et al.*⁵⁴, Leach *et al.*⁴⁹, and Shoyab and Boaze⁵⁰ have co-purified a TPA-binding protein with protein kinase C from brain tissues. There is some evidence that tumour-promoting phorbol esters are intercalated into membranes and directly activate protein kinase C in intact cell systems in a manner similar to that already described for the synthetic diacylglycerol, 1-oleoyl-2-acetyl-glycerol^{41,42,55}. When platelets are stimulated to aggregate and secrete by TPA or PDBu, the 40K protein is as rapidly and heavily phosphorylated as it is by natural extracellular signals such as thrombin. Neither the breakdown of inositol phospholipids nor the formation of endogenous diacylglycerol is observed during the entire period of stimulation. Fingerprint analysis of radiolabelled ³²P-40K protein isolated from the platelets indicates that protein kinase C is responsible for the reaction^{41,42}.

Thus, plausible evidence available to date seems to be compatible with the supposition that protein kinase C is a receptor protein of tumour promoters, and that some, if not all, of the pleiotropic actions of tumour-promoting phorbol esters may be mediated through the action of protein kinase C. This enzyme appears to fulfil the requirements for the proposed receptor which leads to the activation of a variety of cellular functions and proliferation^{6,8}. It is also possible that some other tumour promoters act via protein kinase C. Mezerein, which is classified as a second-stage tumour promoter, can activate protein kinase C⁵⁵. Teleocidin and *Aplysia* toxin, tumour promoters structurally unrelated to phorbol esters, also have the potential to activate protein kinase C *in vitro* at relatively higher concentrations⁵⁶. Although mezerein and the latter classes of tumour promoters do not have diacylglycerol-like structure, it is possible that these compounds may cause analogous changes in membranes.

Synergism with calcium

Although there are some exceptions such as bovine adrenal medullary cells⁵⁷, it is generally the case that, when stimulation of receptors leads to inositol phospholipid breakdown, it simultaneously mobilizes Ca^{2+} . Recently, it has been possible to demonstrate the independent induction of protein kinase C activation and Ca^{2+} mobilization by using the exogenous addition of synthetic diacylglycerol or TPA for the former and a Ca^{2+} ionophore, such as A23187, for the latter^{6,36,37,42,55}. This is shown schematically in Fig. 4. Platelets are particularly suited to the demonstration because the two pathways can be specifically monitored by measuring the phosphorylation of either the 40K or the 20K protein. By contrast, when platelets are stimulated by 'positive signals' such as thrombin, collagen and platelet-activating factor, both the 40K and the 20K proteins are phosphorylated concomitantly²⁷⁻³¹.

In appropriate conditions, it is possible to show that protein kinase C activation and Ca^{2+} mobilization act synergistically to elicit the full physiological response of platelets, such as the release of serotonin^{36,37,42} and lysosomal enzymes⁵⁸. In these experiments the concentration of A23187 (0.2–0.4 μM) is critical since at concentrations greater than 0.5 μM this Ca^{2+} ionophore will itself cause the phosphorylation of 40K protein due to the nonspecific activation of phospholipases and protein kinase C by a large increase in Ca^{2+} concentration. Similarly, it is important to use no more than 30 $\mu\text{g ml}^{-1}$ of synthetic diacylglycerol or 10 ng ml^{-1} TPA, because at higher concentrations (50 $\mu\text{g ml}^{-1}$ or 50 ng ml^{-1} , respectively) they can produce a significant release of platelet constituents without a significant increase in Ca^{2+} concentrations, possibly by acting as membrane fusogens or perturbors (see refs 37,42) and/or by generating superoxide⁵⁹. Recently, Rink *et al.*⁶⁰ have confirmed that at low concentrations neither synthetic diacylglycerol nor TPA *per se*

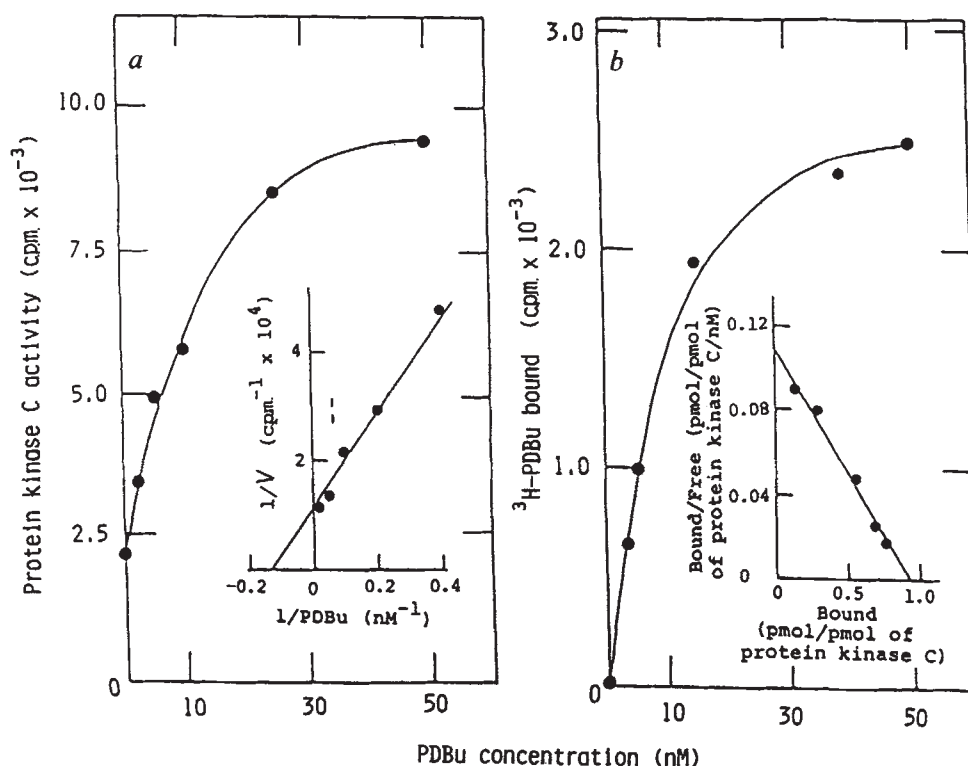


Fig. 3 Effect of PDBu concentration on activation of protein kinase C and on its binding to protein kinase C. *a*, Protein kinase C activity versus PDBu concentration. The protein kinase was assayed at various concentrations of PDBu with 1×10^{-4} M $CaCl_2$ and $8 \mu g ml^{-1}$ phosphatidylserine. Inset, Lineweaver-Burk plots. *b*, 3H -PDBu binding versus its concentration. The binding of 3H -PDBu was assayed in the same conditions by gel filtration on a column of TSK 3000SW, which was attached to a high pressure liquid chromatograph, Toyo Soda Model HLC-803D. Inset, Scatchard plots. The enzyme used was essentially pure. Other details of experimental procedures are given in ref. 43.

mobilizes Ca^{2+} as measured directly with quin 2. The involvement of two synergistic pathways, protein phosphorylation and Ca^{2+} mobilization, may explain the agonist selectivity that is observed for the release reactions from different platelet granules⁵⁸. The balance of these two pathways may exert differential control over different processes within a single activated cell, even if Knight *et al.*⁶¹ have pointed out that there is no difference in their sensitivity to Ca^{2+} mobilization.

The synergistic effects of synthetic diacylglycerol and Ca^{2+} ionophore are not confined to the platelet. For example, when rat mast cells^{6,11} and neutrophils⁵⁸ are stimulated by synthetic diacylglycerol, histamine and lysosomal enzymes such as *N*-acetylglucosaminidase, respectively, are released to some extent. If, however, these cells are incubated with diacylglycerol in the presence of a low concentration of A23187, the release reactions are dramatically enhanced. The Ca^{2+} ionophore alone at the same concentration has little effect. Again, it is plausible that the protein phosphorylation by protein kinase C and the mobilization of Ca^{2+} , both of which are evoked by a single extracellular messenger, are equally indispensable and synergistically effective for causing full physiological responses. Similar synergistic effects have been observed for the proliferation of macrophage depleted peripheral lymphocytes that are treated with phytohaemagglutinin, as measured by the incorporation of radioactive thymidine into DNA⁶². Note, in particular, that tumour-promoting phorbol esters at low concentrations, by inducing the activation of protein kinase C alone, may leave the cell ready to function and proliferate once Ca^{2+} becomes available.

More recently, the synergistic role of protein kinase C and Ca^{2+} mobilization has been extended to several other systems, such as catecholamine release from bovine adrenal medullary cells⁶³, aldosterone secretion from porcine adrenal glomerulosa cells⁶⁴, insulin release from rat pancreatic islets⁶⁵, glycogenolysis in rat hepatocytes^{66,67}, and also concanavalin A-induced activation of bovine lymphocytes⁶⁸. In cell-free systems protein kinase C phosphorylates a large number of proteins in most tissues: for example, vinculin⁶⁹, microtubule-associated proteins¹⁰, ribosomal S6 protein⁷⁰, myosin light chain⁷¹, and initiation factor 2 (β -subunit) for protein synthesis⁷² have been reported to serve as substrates, although the biological significance of the covalent modification of these proteins remains to

be established. A possible cascade reaction from protein kinase C to tyrosine kinases has also been postulated recently, since tyrosyl residues of some endogenous proteins are phosphorylated when cells are treated by TPA or synthetic diacylglycerol⁷³. The substrate proteins for protein kinase C may not be confined to the membrane compartment.

Arachidonic acid cascade

Although one might have anticipated dramatic heterogeneity and variations from tissue to tissue in receptor mechanisms, most tissues seem to possess two major classes of receptors for controlling cellular functions and proliferation (Fig. 4). One class triggers the production of cyclic AMP, while the other induces inositol phospholipid turnover, Ca^{2+} mobilization and frequently arachidonic acid release and cyclic GMP (but not cyclic AMP) production. Thus, phospholipid degradation, Ca^{2+} , arachidonic acid and cyclic GMP appear to be integrated together in a single receptor cascade system, but exactly how is unclear. Michell²⁰ postulated some years ago that inositol phospholipid turnover is involved in Ca^{2+} gate opening. Very recently, Streib *et al.*⁷⁴ have proposed that inositol trisphosphate, which is produced by the signal-induced breakdown of phosphatidylinositol biphosphate, may act as an intracellular mediator of Ca^{2+} mobilization, but this attractive hypothesis is yet to be explored further. Arachidonic acid can be derived from inositol phospholipids through two consecutive reactions, catalysed by phospholipase C and diacylglycerol lipase⁷⁵, and in platelets it can also be derived from phosphatidic acid by its specific lipase⁷⁶ (Fig. 1). However, in most tissues arachidonic acid is also released by phospholipase A₂ from phosphatidylcholine and phosphatidylethanolamine. Perhaps, after receptor stimulation, both phospholipases act in concert.

The role of cyclic GMP, as well as the mechanism by which this cyclic nucleotide is elevated, is still not clearly understood. Although Ca^{2+} sometimes causes direct activation of guanylate cyclase, it is more likely that arachidonic acid peroxide and prostaglandin endoperoxide serve as activators for guanylate cyclase⁷⁷. Cyclic GMP is believed to be an intracellular messenger for various extracellular signals, but it remains a puzzle that cyclic GMP-dependent protein kinase (protein kinase G) shows very similar, if not identical, catalytic properties to those of protein kinase A. Therefore, cyclic GMP may act as

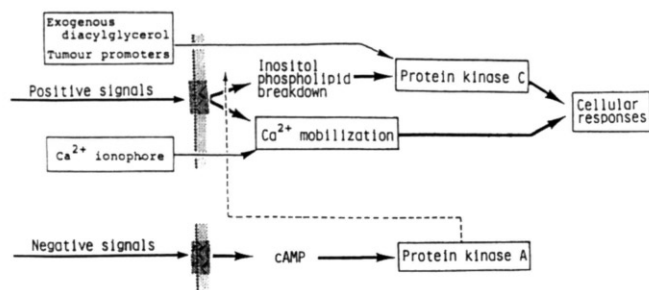


Fig. 4 Synergistic action of protein kinase C activation and Ca^{2+} mobilization for eliciting full cellular responses. In many, but not all, tissues the receptor-linked cascade caused by 'positive signals' is blocked by 'negative signals' which increase cyclic AMP. In some tissues these two receptors do not appear to interact with each other but to function independently⁶.

a negative, rather than a positive, messenger, providing an immediate feedback control that prevents over-response. In support of this proposal is the evidence that, in human platelets stimulated by thrombin, the formation of diacylglycerol from inositol phospholipids, phosphorylation of the 40K protein, and release of serotonin are all inhibited by prior incubation with sodium nitroprusside, which induces cyclic GMP, or with 8-bromo-cyclic GMP⁷⁸. Haslam *et al.*⁷⁹ have suggested that cyclic GMP serves as a negative intracellular messenger on the grounds that sodium nitroprusside is a powerful platelet inhibitor, and Schultz *et al.*⁸⁰ have made a similar proposal for smooth muscle contraction since sodium nitroprusside, nitroglycerol and 8-bromo-cyclic GMP all cause relaxation, although α -adrenergic and muscarinic cholinergic stimulants enhance cyclic GMP levels as well as inositol phospholipid breakdown on contraction. Recent experiments show that cyclic GMP inhibits inositol phospholipid turnover in smooth muscle in a manner similar to that in platelets⁶.

In short, in those cells where cyclic AMP generated by 'negative signals' inhibits the transmission of 'positive signals' (which is not the case in all cells), cyclic AMP and cyclic GMP do not antagonize each other, but similarly inhibit the receptor-linked degradation of inositol phospholipids in membranes, and thereby block cellular functions and proliferation by counteracting the activation of protein kinase C⁶⁻¹¹. These inhibitory actions of cyclic nucleotides probably extend to the mobilization of Ca^{2+} , presumably through activation of protein kinases A and G, respectively^{10,11}.

Finally, it should be emphasized here that the activation of protein kinase C by tumour-promoting phorbol esters is not susceptible to such feedback control by cyclic nucleotides⁴². Unlike diacylglycerol, the tumour promoters may remain in the membrane for prolonged periods of time, since the diterpenes are metabolized only slowly. Perhaps tumour promoters relieve cells of contact inhibition, a process which induces a marked increase in cyclic AMP.

Implications

Although the experimental support reviewed in this article has resulted from studies with a limited number of specific tissues, primarily the platelet, it is plausible that protein kinase C has a crucial role in signal transduction for the activation of many cellular functions and proliferation. With platelets, mast cells, neutrophils and lymphocytes as test cell systems, a possible relationship between Ca^{2+} mobilization and protein phosphorylation catalysed by protein kinase C has been explored. The evidence available thus far suggests that either the increase in Ca^{2+} concentrations or the receptor-linked activation of protein kinase C alone is not sufficient for complete signal transduction, and that their synergistic actions are necessary for the transmembrane control of a variety of cellular events. It is clear that Ca^{2+} and protein kinase C have diverse roles in biological regulation. The precise targets of protein kinase C in each tissue will be clarified by further investigations.

Protein kinase C also appears to be the receptor protein for tumour-promoting phorbol esters. The evidence available to date strongly suggests that some, if not all, of the pleiotropic actions of tumour promoters are mediated through this protein kinase. Further exploration of the roles of protein kinase C may allow an understanding of the biochemical basis of tumour promotion. It is attractive to suggest that the uncontrollable production of an active form of protein kinase C, whether the product of a cellular or of a viral gene, may promote carcinogenesis just like tumour-promoting phorbol esters do.

Protein kinase C may be located on the cross-over point of various pathways in hormone action and cell proliferation, involving Ca^{2+} , inositol phospholipids, arachidonic acid, prostaglandins, and cyclic nucleotides as well as tumour promoters. However the underlying principles and mechanisms outlined here may be modified by the sum of the individual pieces of knowledge that will accumulate rapidly in the next few years.

The investigations reported from this laboratory were supported in part by research grants from the Research Fund of the Ministry of Education, Science and Culture, the Intractable Diseases Division, Public Health Bureau, and the Ministry of Health and Welfare, by a grant-in-aid of New Drug Development from the Ministry of Health and Welfare, and by the Special Coordination Funds from the Science and Technology Agency, Japan.

- Inoue, M., Kishimoto, A., Takai, Y. & Nishizuka, Y. *J. biol. Chem.* **252**, 7610-7616 (1977).
- Takai, Y. *et al. J. biol. Chem.* **254**, 3692-3695 (1979).
- Takai, Y., Kishimoto, A., Kikkawa, U., Mori, T. & Nishizuka, Y. *Biochem. biophys. Res. Commun.* **91**, 1218-1224 (1979).
- Kishimoto, A., Takai, Y., Mori, T., Kikkawa, U. & Nishizuka, Y. *J. biol. Chem.* **255**, 2273-2276 (1980).
- Nishizuka, Y. *Molec. Biol. Biochem. Biophys.* **32**, 113-135 (1980).
- Nishizuka, Y. *Phil. Trans. R. Soc. B.302*, 101-112 (1983).
- Nishizuka, Y. & Takai, Y. *Cold Spring Harb. Conf. Cell Proliferation* **8**, 237-249 (1981).
- Nishizuka, Y. *Trends biochem. Sci.* **8**, 13-16 (1983).
- Takai, Y., Kishimoto, A. & Nishizuka, Y. in *Calcium and Cell Function* Vol. 2 (ed. Cheung, W. Y.) 385-412 (Academic, New York, 1982).
- Takai, Y., Kikkawa, U., Kaibuchi, K. & Nishizuka, Y. *Adv. Cyclic Nucleotide Res.* (in the press).
- Nishizuka, Y., Takai, Y., Kishimoto, A., Kikkawa, U. & Kaibuchi, K. *Recent Prog. Horm. Res.* **40**, 301-345 (1984).
- Kaibuchi, K., Takai, Y. & Nishizuka, Y. *J. biol. Chem.* **256**, 7146-7149 (1981).
- Mori, T., Takai, Y., Minakuchi, R., Yu, B. & Nishizuka, Y. *J. biol. Chem.* **255**, 8378-8380 (1980).
- Wise, B. C. *et al. J. biol. Chem.* **257**, 8489-8495 (1982).
- Kuo, J. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7039-7043 (1980).
- Kikkawa, U., Takai, Y., Minakuchi, R., Inohara, S. & Nishizuka, Y. *J. biol. Chem.* **257**, 13341-13348 (1982).
- Kishimoto, A., Kikkawa, N., Shiota, M. & Nishizuka, Y. *J. biol. Chem.* **258**, 1156-1164 (1983).
- Nimmo, H. G. & Cohen, P. *Adv. Cyclic Nucleotide Res.* **8**, 145-266 (1977).
- Krebs, E. G. & Beavo, J. A. *Rev. Biochem. Chem.* **48**, 923-959 (1979).
- Michell, R. H. *Biochim. biophys. Acta* **415**, 81-147 (1975).
- Berridge, M. J. *Molec. cell. Endocr.* **24**, 115-140 (1981).
- Michell, R. H., Kirk, C. J., Jones, L. M., Downes, D. P. & Creba, J. A. *Phil. Trans. R. Soc. B.296*, 123-137 (1981).
- Hokin, M. R. & Hokin, L. E. *J. biol. Chem.* **203**, 967-977 (1953).
- Downes, P. & Michell, R. H. *Cell Calcium* **3**, 467-502 (1982).
- Berridge, M. J., Dawson, R. M. C., Downes, C. P., Heslop, J. P. & Irvine, R. F. *Biochem. J.* **112**, 473-482 (1983).
- Abdel-Latif, A. A., Akhtar, R. A. & Hawthorne, J. N. *Biochem. J.* **162**, 61-73 (1977).
- Kawahara, Y., Takai, Y., Minakuchi, R., Sano, K. & Nishizuka, Y. *Biochem. biophys. Res. Commun.* **97**, 309-317 (1980).
- Sano, K., Takai, Y., Yamanishi, J. & Nishizuka, Y. *J. biol. Chem.* **258**, 2010-2013 (1983).
- Lyons, R. M., Stanford, N. & Majerus, P. W. *J. clin. Invest.* **56**, 924-936 (1975).
- Haslam, R. J. & Lynham, J. A. *Biochem. biophys. Res. Commun.* **77**, 714-722 (1977).
- Ieyasu, H., Takai, Y., Kaibuchi, K., Sawamura, M. & Nishizuka, Y. *Biochem. biophys. Res. Commun.* **108**, 1701-1708 (1982).
- Yagi, K., Yazawa, M., Kakiuchi, S., Ohshima, M. & Uenishi, K. *J. biol. Chem.* **253**, 1338-1340 (1978).
- Imaoka, T., Lynham, J. A. & Haslam, R. J. *J. biol. Chem.* **258**, 11404-11414 (1983).
- Rittenhouse-Simmons, S. *J. clin. Invest.* **63**, 580-587 (1979).
- Bell, R. L. & Majerus, P. W. *J. biol. Chem.* **255**, 1790-1792 (1980).
- Kaibuchi, K., Sano, K., Hoshijima, M., Takai, Y. & Nishizuka, Y. *Cell Calcium* **3**, 323-335 (1982).
- Kaibuchi, K. *et al. J. biol. Chem.* **258**, 6701-6704 (1983).
- Hecker, E. *Meth. Cancer Res.* **6**, 439-484 (1971).
- Weinstein, I. B., Lee, L. S., Mufson, A. & Yamasaki, H. *J. supramolec. Struct.* **12**, 195-208 (1979).
- Blumberg, P. M. *CRC Crit. Rev. Tox.* **8**, 153-197 (1980).
- Castagna, M. *et al. J. biol. Chem.* **257**, 7847-7851 (1982).
- Yamanishi, J. *et al. Biochem. biophys. Res. Commun.* **112**, 778-786 (1983).
- Kikkawa, U., Takai, Y., Tanaka, Y., Miyake, R. & Nishizuka, Y. *J. biol. Chem.* **258**, 11442-11445 (1983).
- Dunphy, W. G., Kochenburger, R. J., Castagna, M. & Blumberg, P. M. *Cancer Res.* **41**, 2640-2647 (1981).
- Horowitz, A. D., Greenebaum, E. & Weinstein, I. B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2315-2319 (1981).
- Solanki, V. & Slaga, T. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2549-2553 (1981).

47. Ashendel, C. L., Staller, J. M. & Boutwell, R. K. *Biochem. biophys. Res. Commun.* **111**, 340–345 (1983).
48. Sando, J. J. & Young, M. C. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2642–2646 (1983).
49. Leach, K. L., James, M. L. & Blumberg, P. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4208–4212 (1983).
50. Shoyab, M. & Boaze, R. Jr *Cold Spring Harb. Conf. Cell Proliferation* **11** (in the press).
51. Delclos, K. B., Yeh, E. & Blumberg, P. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3054–3058 (1983).
52. Kraft, A. S., Anderson, W. B., Cooper, H. L. & Sando, J. J. *J. biol. Chem.* **257**, 13193–13196 (1982).
53. Shoyab, M. & Todaro, G. J. *Nature* **388**, 451–455 (1980).
54. Niedel, J. E., Kuhn, L. J. & Vandenbark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **80**, 36–40 (1983).
55. Kikkawa, U., Miyake, R., Tanaka, Y., Takai, Y. & Nishizuka, Y. *Cold Spring Harb. Conf. Cell Proliferation* **11**, (in the press).
56. Fujiki, H. *et al.* in *Cellular Interactions by Environmental Tumor Promoters* (ed. Sugimura, T.) (Japan Sci. Soc., Tokyo, in the press).
57. Swilen, A. M. F., Hawthorne, J. N. & Azia, N. *Biochem. Pharmac.* **32** (in the press).
58. Kajikawa, N. *et al.* *Biochem. biophys. Res. Commun.* **116**, 743–750 (1983).
59. Witz, G., Goldstein, B. D., Amoroso, M., Stone, D. S. & Troll, W. *Biochem. biophys. Res. Commun.* **97**, 883–888 (1980).
60. Rink, T. J., Sanchez, A. & Hallam, T. J. *Nature* **305**, 317–319 (1983).
61. Knight, D. E., Hallam, T. J. & Scrutton, M. C. *Nature* **296**, 256–257 (1982).
62. Kaibuchi, K., Sawamura, M., Kikkawa, U., Takai, Y. & Nishizuka, Y. in *Inositol and Phosphoinositides* (eds Bleasdale, J. E., Eichberg, J. & Hauser, J.) (Humana, Dallas, in the press).
63. Knight, D. E. & Baker, P. F. *FEBS Lett.* **160**, 98–100 (1983).
64. Kojima, I., Lippes, H., Kojima, K. & Rasmussen, H. *Biochem. biophys. Res. Commun.* **116**, 555–562 (1983).
65. Zawulich, W., Brown, C. & Rasmussen, H. *Biochem. biophys. Res. Commun.* **117**, 448–455 (1983).
66. Garrison, J. C., Johnsen, D. E. & Campanile, C. P. *J. biol. Chem.* (in the press).
67. Fain, J. N., Li, S.-Y., Litosch, I. & Wallace, M. *Biochem. biophys. Res. Commun.* (in the press).
68. Mastro, A. M. & Smith, M. C. *J. cell. Physiol.* **116**, 51–56 (1983).
69. Werth, D. K., Niedel, J. E. & Pastan, I. *J. biol. Chem.* **258**, 11423–11426 (1983).
70. Le Peuch, C. J., Ballester, R. & Rosen, O. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6858–6862 (1983).
71. Naka, M., Nishikawa, M., Adelstein, R. S. & Hidaka, H. *Nature* **306**, 490–492 (1983).
72. Schatzman, R. C., Griño, J. A., Merrick, W. C. & Kuo, J. F. *FEBS Lett.* **159**, 167–170 (1983).
73. Martin, G. S. & Gilmore, T. *Nature* **306**, 487–490 (1983).
74. Streb, H., Irvine, R. F., Berridge, M. J. & Schulz, I. *Nature* **306**, 67–69 (1983).
75. Bell, R. L., Kennerley, D. A., Stanford, N. & Majerus, P. W. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3238–3241 (1979).
76. Billah, M. M., Lapetina, E. G. & Cuatrecasas, P. *J. biol. Chem.* **256**, 5399–5403 (1981).
77. Graff, G., Stephenson, J. H., Glass, D. B., Haddox, M. K. & Goldberg, N. D. *J. biol. Chem.* **253**, 7662–7676 (1978).
78. Takai, Y., Kaibuchi, K., Matsubara, T. & Nishizuka, Y. *Biochem. biophys. Res. Commun.* **101**, 61–67 (1981).
79. Haslam, R. J., Salama, S. E., Fox, J. E. B., Lynham, J. A. & Davidson, M. M. L. in *Platelets: Cellular Response Mechanisms and their Biological Significance* (eds Rotman, A., Meyer, F. A., Gitler, C. & Silberberg, A.) 213–231 (Wiley, New York, 1980).
80. Schultz, K. D., Schultz, K. & Schultz, G. *Nature* **265**, 750–751 (1977).

ARTICLES

Dynamics of the oceanic general circulation: the potential vorticity field

W. R. Holland*, T. Keffer† & P. B. Rhines†

* National Center for Atmospheric Research, Boulder, Colorado 80307, USA

† Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

The general circulation of the oceans is determined by vorticity, temperature and salinity. The theory of large-scale circulation puts a strong constraint on the fields of potential vorticity, q . New calculations and observations of the North Atlantic and North Pacific, using the field of potential vorticity as a point of comparison, reveal strong similarities. In particular, the observations can be modelled by gyres around which geostrophic contours deform to follow the circulation and within which q is uniform. The effects of injected waters with relatively low potential vorticity are significant.

THE calculation of the general circulation of the major oceans is notoriously difficult even though it is a simple matter to write down the dynamical equations that describe the motion of the elements within a relatively thin layer on the surface of the Earth rotating with angular velocity Ω . Mathematically, the difficulty is that the equations that in principle determine the motion within the oceans in terms of the boundary conditions (at the margins and at the surface) are non-linear, in some connections to a high order. Our objective in what follows is to show that it is nevertheless possible to construct a model of the wind-driven circulation in major oceans which yields results in qualitative agreement with observation.

In the major oceans, the general character of the circulation is easily described. The uppermost 1,000 m or so of the oceans tends to be dominated by great circulatory gyres, one or two to an ocean basin, which are generated by wind stress at the surface. At greater depths, however, the pattern of flow is dominated by differences of the temperature and salinity of different water masses. One consequence of such thermohaline circulations is to create the vertical stratification of density in the oceans—or, more accurately, the stratification of potential density ρ defined as the density of the fluid element concerned after the pressure is adjusted adiabatically to some reference value.

To construct a theory of the circulation of the oceanic gyres, we exploit the vast difference between the response time of the wind-driven circulation (of the order of 10 yr) and that of the thermohaline circulation (of the order of 1,000 yr). From this point of view, the wind-driven circulation may be thought to ride on an unchanging basic stratification of ρ which is determined by the long-term (1,000 yr) thermohaline circulation. Even so, it is central to the argument which follows that surfaces

of constant ρ which are at great depth at middle latitudes rise with increasing latitude and eventually intersect the sea surface along 'outcrop' curves. Since surface water can move downward along surfaces of constant ρ without working against gravity, the outcrop curves are 'windows' by means of which water marked by atmospheric interactions can reach great depths in the oceans (Fig. 1). Hence the sea surface is a key site where the thermodynamic and mechanical interactions with the atmosphere take place and from where the memory of these interactions is carried with the fluid down into the ocean's interior.

The stable stratification of density in the oceans constrains fluid motions to lie nearly along surfaces of constant potential density, known as isopycnal surfaces. A further constraint acting on the motion of a fluid element within an isopycnal surface arises from the rotation of the Earth. Specifically, in the context of planetary fluids, if $\mathbf{u}$ is the velocity field of the flow (referred to the surface of the rotating Earth) where frictional and externally applied forces are negligible and $\nabla \times \mathbf{u}$ is the relative vorticity field, then the potential vorticity, defined as $q = (2\Omega + \nabla \times \mathbf{u}) \cdot \nabla \rho / \rho$ is, like the potential density, a conservative quantity in the sense that q for each fluid element remains constant following the motion. Accordingly, the directions of flow on surfaces of constant ρ are contours of constant q , known as geostrophic contours or isostrophs. In other words, a theory of the scalar fields q and ρ is tantamount to a theory of the general circulation.

Theoretical review

The mathematical problem of predicting both the circulation and the basic density field is strongly nonlinear, even in its simplest form. The equations describe both the determination

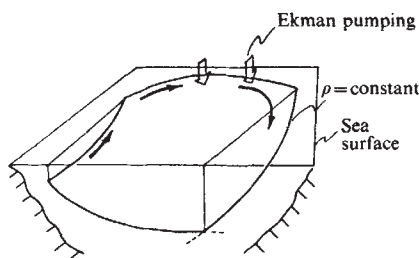


Fig. 1 Schematic diagram of the injection of surface waters downward into the wind-driven ocean gyre. The inflowing water is squeezed into a narrow band near the eastern boundary, joining the more massive stream emanating from the western boundary. Mixing rates near the 'outcrop' curves determine how rapidly the newly injected waters invade the core of the gyre.

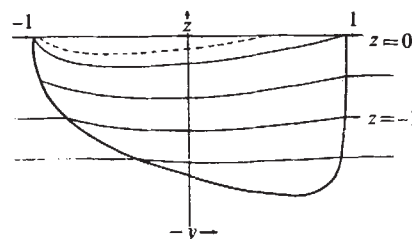


Fig. 2 A meridional section through a theoretical model of the subtropical wind-driven circulation¹. The thickness between constant-density surfaces (light full lines) increases poleward, such that the potential vorticity is uniform within the bounds of the gyre (heavy lines). The gyre centres recede poleward with increasing depth, and the poleward side of the gyre is steep.

of q and ρ in the interior by boundary values 'upstream' at the sea surface, and also stirring and mixing processes in the deep interior which, given time, cause major alteration of these values. Even if we could solve the interior problem generally, there are regions with dynamics of higher order which interact essentially with the interior: particularly, the boundary layers at the sides, top and bottom of the oceans.

The equations governing the time-averaged (symbols with an overbar) flow can, in the presence of a conservative scalar like potential density, be reduced to a conservation principle for a dynamically active scalar, the potential vorticity q . This is

$$\begin{aligned}\bar{\mathbf{u}} \cdot \nabla \bar{q} &= -\nabla \cdot \bar{\mathbf{u}}' \bar{q}' + \bar{F} - \bar{D} \\ \bar{\mathbf{u}} \cdot \nabla \bar{\rho} &= -\nabla \cdot \bar{\mathbf{u}}' \bar{\rho}' + \bar{H}\end{aligned}\quad (1)$$

Here F is the external forcing (by winds, and air-sea buoyancy transport), D is the effect of small-scale dissipation and H is external forcing of buoyancy, plus the effect of small-scale mixing. The local-vertical component of 2Ω is the Coriolis frequency, $f = f_0 + \beta y$. x , y , z are east north and upward local cartesian coordinates. Perturbations about the mean are denoted by primes. For quasi-geostrophic baroclinic ocean dynamics with lateral scale much larger than the Rossby deformation radius (~ 50 km) the relative vorticity is much smaller than the 'stretching' term. Then

$$q = (f/\rho)(\partial\rho/\partial z)$$

and the potential vorticity can be evaluated from hydrographic measurements of ρ alone.

The free Rossby-wave modes of the large-scale ocean, instabilities of the mean circulation and geostrophic turbulence all involve interplay between the mean and perturbation fields of q . These transient currents provide eddy transports appearing on the right-hand side of equation (1). Where these eddy and dissipation terms are weak we have as a first approximation

$$\begin{aligned}\bar{\mathbf{u}} \cdot \nabla \bar{q} &= 0 \\ \bar{\mathbf{u}} \cdot \nabla \bar{\rho} &= 0\end{aligned}\quad (2)$$

for the circulation of the interior ocean. These equations were introduced in the pioneering work of Welander². Equations (2) are, however, underdetermined in a large domain of the fluid. The right-hand side terms of equations (1) must arise to select among possible laminar solutions. We have added the interior eddy stresses which in this way select a unique solution.

The large-scale potential vorticity is proportional to f/dh , where dh is the infinitesimal thickness between neighbouring constant- ρ surfaces. Its conservation expresses the rigidity

imparted to a fluid by planetary rotation, for equations (2) state merely that an infinitesimal line of fluid particles moves such as to keep constant the projection of the rotation vector Ω on it.

Despite the simplicity of these relationships (and some early awareness of the importance of these geostrophic contours^{3,4}), the analysis of the general circulation using maps of constant- ρ surfaces is relatively recent (being exploited for the analysis of salinity, oxygen and nutrient fields particularly by Montgomery⁵ and Reid⁶). Analysis of geostrophic contours ($q = \text{constant}$ on surfaces of $\rho = \text{constant}$) is even more recent⁷⁻⁹.

Predicted nature of the potential vorticity field. In much of the oceans the classic Sverdrup relation $\beta \int_{-H}^0 v \, dz = fw_0$, where w_0 is the wind-forced vertical velocity at the base of the upper boundary layer, follows from equation (2). If this net transport of mass, a known function of the wind-stress, were to be distributed over too great a depth interval, q would not be very different from its 'rest' value, proportional to βy . Blockage of the geostrophic contours by continental boundaries would then prevent flow in most of the body of the ocean, invalidating this as a possible solution. Linearized 'spin-up' models of the ocean circulation for this reason evolve towards a singular velocity distribution, with the circulation confined to a thin layer at the sea surface, one where the direct effect of the wind is not shielded by the stratification.

The true circulation is essentially nonlinear, in the sense that flow-induced modification of the geostrophic contours is necessary if the interior ocean is to move meridionally. If the intensity of the circulation were great enough, one could imagine 'islands' of q rising above the sloping rest state, $q = \text{constant} \times \beta y$. There, fluid might circulate about an entire gyre without experiencing external stresses or suffering dissipation near the western boundary. In the vertically integrated sense, dissipation must still occur to balance the integrated wind curl, but this could occur in isolated depth intervals, leaving the western boundary current largely inertial. Regardless of the detailed boundary layer dynamics, the mid-ocean circulation must cause, through the geostrophic tilt of its ρ surfaces, the geostrophic contours to make major north-south excursions. The only alternative would be massive diffusion of momentum or density.

In a recent theory of the circulation^{1,10} based on equations (1) with weak right-hand sides, these islands of q are found to be essential. The geostrophic contours that close on themselves without touching the outcrop 'window' provide easy passage for the fluid, without the necessity for strong dissipation. The (known) vertically integrated velocity is distributed in the vertical as eddy 'wave-drag' carries momentum from one level to the next. The eddies themselves are spawned by instability of the general circulation, giving this dynamical evolution a strong feedback.

For a given Sverdrup transport the 'burrowing' of the circulation downward can extend only so far. If major distortions of

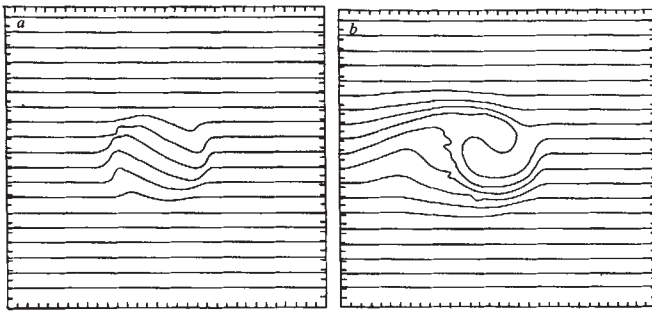


Fig. 3 The spin-up of the circulation from rest, when the winds arise impulsively¹³. This is a two-layer model without lateral boundaries, in which eddy effects are parameterized as a diffusion of potential vorticity. The shear of the mean flow augments the diffusion, and low values of q 'leak' to the exterior. This transient field then propagates westward as a baroclinic Rossby wave. *a* and *b* show $q(x, y)$ during the first revolution of the gyre.

the rest-state q field occur at each level, a monotonically decreasing flow will have a scale-depth given by $D \sim (f/N)\sqrt{U/\beta}$ where U is the typical magnitude of the wind-driven current, and N is the buoyancy frequency. This follows simply by requiring that the density-tilt contribution to ∇q has the same magnitude as β so that the geostrophic contours can indeed depart from latitude circles. The process can continue right to the sea floor in the most intense parts of the gyre¹¹.

The actual distribution of q predicted by this theory depends on the details of eddy-induced dissipation in the western boundary current. If that dissipation is sufficiently weak, or conversely so strong that the boundary current acts as a 'mixing valve', the theory shows that q becomes homogenized to a single value within the closed streamlines of the gyre: the islands of q become plateaus. Such creation of regions of uniform potential vorticity is predicted to be an important feature of both atmospheric and oceanic circulations. It has analogues in other flows with closed streamlines, for example in the expulsion of magnetic field lines from fluid convection cells in the solar photosphere¹².

Typical theoretical solutions following this scheme (Fig. 2) exhibit a poleward migration of the subtropical gyre with increasing depth, and at its poleward side the gyre is steeply terminated. The potential vorticity develops homogeneous ledges which allow free passage of fluid across great extremes of latitude.

Evolution from a state of rest. The evolution of the potential vorticity field from an initial state of rest (Fig. 3) involves shear-augmented diffusion of potential vorticity¹³, which provides an enhanced diffusion of q along mean streamlines. This two-layer numerical simulation¹⁴ is the solution of the time-dependent large-scale equations with the role of eddies simulated by vertical and lateral friction. For simplicity, there are

no lateral boundaries, and the wind-stress curl varies linearly in x inside a circle of radius a . (The oceanic problem corresponds to the right half of this flow, with a coastal boundary lying north-south along the midline, and a western boundary current replacing the left half-plane.) When the winds are switched on, the initial Sverdrup circulation is uniform with depth (the barotropic adjustment is instantaneous in these equations), and the q field is just proportional to βy . The streamlines are circular.

The lower-layer field of q (shown in Fig. 3) is wrapped round by the circulation, developing large gradient in the radial direction. This stretching of the q contours greatly increases the diffusion of q . The systematic westward propagation of transients by baroclinic Rossby waves redistributes the potential vorticity that 'leaks' out of the gyre, while radiating energy westward to infinity. After the transient phase we are left with a steady circulation characterized by a plateau of homogenized q . As predicted in the theory, the value of the constant q is not the average value of the initial conditions, but rather the value at the poleward extremity of the gyre (According to classical linear theory the time required to establish such a circulation would be the transit time L/c , for baroclinic Rossby waves to cross the gyre (L is the breadth of the gyre and c the long-wave speed $\sim \beta(NH/f_0)^{1/2}$, where h is the depth). The present theory, however, predicts the adjustment time to be intermediate between the circulation time, L/U , and the diffusion time, L^2/κ , where κ is the lateral diffusivity of potential vorticity^{13,14}.)

Points of entry into the circulation. Fluid enters the circulation of the subtropical gyre from the upper mixed layer at the outcrop lines (Fig. 1). Once it descends out of reach of the atmosphere, subsequent modification of q is thought to be gradual. Further eddy 'scattering' processes are active, however, in preventing the cyclic travel about the gyre from being perfectly periodic. These eddy transports largely account for the creation, mixing and destruction of water masses. There is thus a competition in the determination of the field of q between the propagation of sea-surface boundary values along geostrophic contours and (as in the theory described above) the determination of q by eddy momentum flux across q contours and ρ surfaces. The classical non-diffusive thermocline theory, and particularly recent applications to the wind gyre^{2,15}, centre on the latter more hyperbolic solution scheme, essentially integrating downstream and downward from the outcrop lines. Thus, lying between the recirculating gyre and the quiescent 'shadow zone' of the quasi-geostrophic theory there is a band of directly injected fluid with memory of upstream boundary conditions at the sea surface. In certain density intervals the injection rate is so great that q , being proportional to $\partial\rho/\partial z$, takes a relative minimum; these are termed 'mode' waters¹⁶.

The basic planetary rotation restricts the size of this band of directly injected waters. It typically occupies a small fraction $\sim L_0/a$ of the width of the ocean, where L_0 is the north-south extent of the gyre and a is the Earth's radius. Even though the

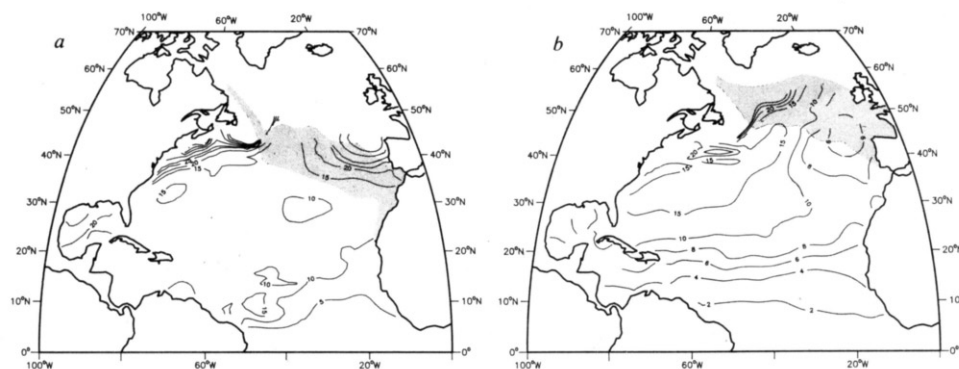


Fig. 4 The potential vorticity of the constant potential density layer $\rho = 26.5\text{--}27.0\text{‰}$ (*a*) and $\rho = 27.0\text{--}27.3\text{‰}$ (*b*) in the North Atlantic ($10^{-13} \text{ cm}^{-1} \text{ s}^{-1}$). The upper layer exhibits nearly homogeneous q within the circulation gyre, with a steep discontinuity at its equatorward edge. The lower layer, at the base of the wind gyre, shows evidence of flow pathways connecting the sea surface at high latitude (shaded) with the deep thermocline levels in the subtropics. The low values of q in the east correspond to southward flowing mode waters from the subpolar gyre¹⁶. The large gradient of q in the outcrop regions (shaded) is a seasonal effect.

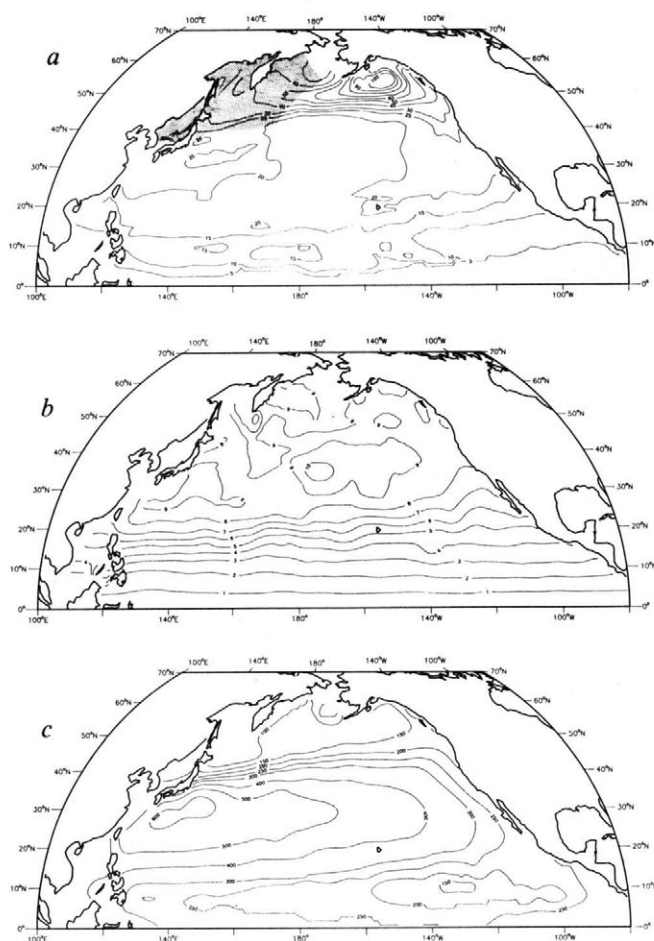


Fig. 5 As in Fig. 4, but for the North Pacific. *c*, Shows in addition the depth (in metres) of the surface mapped in *a*. Again the wind-driven subtropical gyre is homogenized in potential vorticity across its entire span. A strong gradient separates this from the northern (subpolar) gyre at the upper level (*a*). A tongue of high values of q enters the gyre on its eastern side. It is a shallow thermohaline circulation driven from the north, yet is guided by the q and u fields of the wind gyre, with which it mingles. The disappearance of the 'tongue' is clear evidence of the mixing of potential vorticity laterally or vertically by eddies. The deeper layer (*b*) is virtually homogenized across both of the principal circulation gyres, in the absence of the strong thermohaline circulation seen in the North Atlantic. Small variations of q remaining in the gyre are reliable features, which may provide constraints on the flow.

source regions for these water masses may be extensive, the theory suggests that they are pressed into narrow streams on the eastern sides of the oceans (a/L_0 has been called the 'recirculation index', for it measures the amount of 'round-and-round' flow in the gyre, relative to the amount injected from the upper boundary layer. In this circumstance the newly entering fluid is joined by a greater stream emanating from the western boundary current. Entry of both chemical tracers and q thus resembles a forced convection problem, in which mixing between newly entering and older waters determines the distributions deep in the interior. An interesting new numerical model that exhibits all these effects in the context of a steady circulation with strong diffusion but no eddies has been developed by Cox and Bryan¹⁷. The choice of model configuration (mixing rates, layer depths) in the vicinity of the outcrop lines is crucial to the global solution. This simple result of geometry and 'geostrophy' elevates the role of the western boundary current, which joins the lesser flux of newly injected waters. It suggests that a scheme for calculating the circulation by downstream integration from the outcrop boundaries is incomplete, and that major parts

of the circulation gyres may be sufficiently isolated to develop their q fields by internal adjustment.

There are branches of the general circulation, particularly those for which $a/L_0 \sim 1$, that will more resemble 'source-sink' flows with open q contours, than recirculating gyres. The thermohaline flow beneath the wind gyre may in addition rely on the deformed bowl-like density surfaces overhead, or the sloping bottom topography below, to create paths favourable to meridional flow in mid-ocean.

Oceanographic observations

Earlier works^{8,9} described the potential vorticity field of the North Atlantic. Here we concentrate on two specific density levels and compare observations of the North Atlantic with those of the North Pacific, and with theory and numerical experiments. The data set is the 1°-global time-averaged ensemble of Levitus¹⁸, with 5° smoothing. Comparison with more refined sections and surveys⁸ has shown the essential features to be reproducible. Boundary currents and other meso-scale features are smoothed over here, but they require special treatment in any case, owing to the lack of relative-vorticity data.

The two intervals are $\rho = 26.5\text{--}27.0\%$, corresponding to the middle depth of the subtropical wind gyres, and $\rho = 27.0\text{--}27.3\%$, which corresponds to the base of the wind gyre. Potential vorticity is calculated as $f\delta\rho/\delta z$ where $\delta\rho$ is the surface-referenced potential density increment and δz is the vertical distance between top and bottom levels of that increment.

Mid-depth in the wind-gyre. Figure 4 shows the potential vorticity field for these two surfaces in the North Atlantic. The shallower (26.5–27.0) layer (Fig. 4a) lies just below the deepest penetration of wintertime convection, except at its outcrop line near 40° N. The first impression is that, indeed, the q distribution is greatly altered by the circulation. Furthermore, it is here that the conditions for expulsion of gradients of q seem best to be met and, indeed, the gyre is dominated by nearly uniform potential vorticity across its entire width. In a state of rest, the potential vorticity βy would more than double across this span of latitude, but the circulation-induced deformation of density surfaces almost exactly cancels this gradient. There are smaller systematic variations of q within the gyre⁸, perhaps representing the 'coloration' of the gyre by the western boundary current and deep convection.

At this level the recirculation index, described above, is moderately large (3–5, say). Consistent with this emphasis on the round-and-round nature of the circulation is a recent radiochemical analysis of Sarmiento¹⁹ who finds that the amount of tritium entering the gyre is too great by a factor of 3–5 to be accounted for by simple Ekman pumping (that is, the Ekman velocity multiplied by the observed surface tritium concentration, integrated over the outcrop 'window'). This suggests that lateral mixing processes, and possibly cross-frontal diapycnal mixing, are strong, and indeed overwhelm the laminar picture of gyre ventilation. As we remarked earlier, the most likely region for the invasion of the gyre to occur is where the emanating western boundary current 'grazes' the upper surface. Once injected into the gyre, the fluid makes several circuits around the gyre before exiting. Its typical north-south velocity greatly exceeds the velocity which would be required by mass-flow continuity alone (An independent measure of the recirculation index is the ratio of observed transport (say, 30 Sv at the Florida Straits and 80 Sv at Cape Hatteras; 1 Sv = $10^6 \text{ m}^3 \text{ s}^{-1}$) to the downward Ekman flux between $\rho = 26.2$ and 27.4 outcrops, estimated to be 8 Sv¹⁹. This ratio is thus about 4 for the Sverdrup gyre and about 10 for the inner, intense recirculation in the western North Atlantic.)

The corresponding (slightly lighter) density surface in the North Pacific (Fig. 5a, c) again has a maximum depth of 500–600 m. In this much larger ocean we see even more dramatic homogenization of potential vorticity throughout most of the subtropical gyre, between 15° N and 40° N latitude. The region of homogenization is shown in section in Fig. 6, in which the

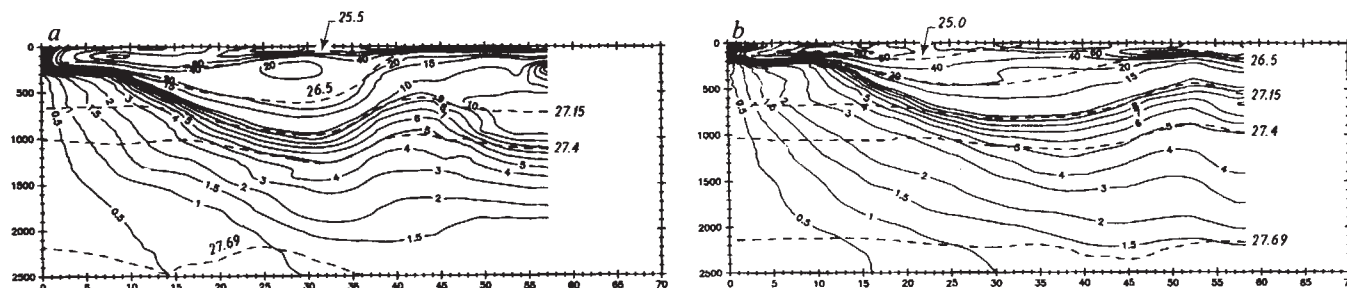


Fig. 6 Meridional (latitude versus depth) sections of the potential vorticity (solid) and potential density (dashed) in the North Pacific. *a*, 150° E longitude; *b*, 150° W longitude. The polar front weakens noticeably as one moves eastwards. The wind-driven circulation reaches at least 1,500 m depth, and involves extensive homogenization of q on the constant- ρ surfaces. The greater fraction of constant- ρ surfaces fails to make contact with the sea surface anywhere in this ocean.

dip and hump in the isopycnal surfaces clearly outline the subtropical and subpolar gyres, respectively. At 150° E longitude this 'bubble' extends to perhaps 1,500 m depth, in accord with refined estimates of the penetration depth¹, and in agreement with Coats²⁰ analysis of two particular density sections.

There are, however, important differences between the two oceans. In the North Atlantic the entire subpolar gyre above $\rho = 27.2$ is seasonal. That is, winter surface densities are generally greater than this value. Consequently, the subpolar gyre is missing from Fig. 4*a*, and the outcrop window lies at the front between the two gyres. By contrast, the analogous surface density in the North Pacific has far less exposure to the atmosphere. The outcrop region stippled on Fig. 5*a* is confined to a relatively small region in the north-west. Because the boundary between the gyres lies along the dividing line between upward and downward Ekman pumping, it is clear that no simple, direct ventilation of this density surface is possible, by the mechanism of wind-driving. It remains for vertical mixing and lateral eddy processes, of the type envisaged in the homogenization theory and the numerical simulations, to bring the potential vorticity of the gyre to equilibrium.

Particularly relevant to ventilation mechanisms is a narrow tongue of high-potential vorticity water entering the subtropical gyre on its northeastern side, at 100–400 m depth. The narrowness of this flow is consistent with the dynamical arguments above, which would have the entering fluid pressed up against the eastern boundary. In this case the tongue is a part of the rather weak, shallow thermohaline mode of the Pacific in which fresh, highly oxygenated waters from the far north penetrate equatorward. (Salinity and oxygen maps⁶ are consistent with our dynamical maps.) This mode of circulation is clearly affected by the wind gyre, and may, in fact act as a small perturbation to it, flowing along isostrophs determined principally by the wind circulation.

The largeness of q found in the subpolar gyre is a distinctly thermohaline aspect of the North Pacific. A light, fresh layer of water lies in the top 150 m, due to the excess of precipitation over evaporation (the evaporation being noticeably less at comparable latitudes in the Pacific than in the Atlantic, because of the warmer surface waters of the Atlantic). The density surface shown here lies at the base of this layer, in a sharp halocline. The $f\partial\rho/\partial z$ values are correspondingly high.

This precipitation and attendant vertical mixing act as a strong source of potential vorticity in the northern gyre of the North Pacific. Potential vorticity is consequently not homogenized in the northern gyre, for the 'source-free' assumption of the Prandtl–Batchelor theorem is violated. Also, the high values of the northern gyre q are not efficiently communicated to the subtropical gyre lying to the south; the 'impedance' of the gyre boundary is too great, eddy mixing and persistent thermohaline circulation causing only a weak tongue of large q to penetrate to the south. Other chemical tracers show this same contrast between gyres.

The base of the wind gyre. Contrast this level, at roughly mid-depth in the subtropical wind gyre, with the deeper surface

shown in Figs 4*b* and 5*b*. In the North Atlantic this range of ρ (27.0–27.3) reaches the sea surface near the boundary between subpolar and subtropical gyres, and extends down to 800 m depth in the core of the recirculation. Despite the weakened or non-existent Ekman pumping, both potential vorticity and passive chemical tracers succeed in flowing along north–south running paths, created in part by the presence of the wind gyre overhead. Evidence from tracers suggests principally southward flow, although at the next level below, a northward return may occur at the eastern side of the ocean to feed the Norwegian/Greenland Sea sinking regions²¹. The great penetration southward along these geostrophic contours (indeed, so much so as to produce small potential vorticity 'mode waters'¹⁶) again emphasizes that the principal ventilation mechanism is not simple laminar Ekman pumping.

The corresponding surface density in the North Pacific, by contrast, is everywhere shielded from direct influence of the atmosphere (ranging from 500 m depth in the subpolar, northern gyre to 900 m depth in the core of the subtropical gyre). Yet, paradoxically, the wind gyres are far stronger and more extensive at this depth than in the North Atlantic. The subtropical circulation cannot be driven by the 'direct ventilation' mechanism here, or even on the shallower surface described above. Eddy stresses are, instead, a prime candidate. With the sources and sinks of potential vorticity correspondingly weak, lateral eddy stirring is apparently capable of mixing efficiently across the gyre boundary (as well as maintaining the vertical structure of the gyres). Both subpolar and subtropical gyres equilibrate at a single, uniform value of q , extending over 25°–55° N latitude, and across the full 10,000 km breadth of the ocean. The great span in both vertical and horizontal of this massive homogenization is evident in the meridional sections (Fig. 6). In the westernmost section (at 150° E longitude) there is an isolated minimum of potential vorticity (the 20–contour) corresponding to the mode water, analogous to the 18 °C water of the Atlantic¹⁶. The spatial extent of such thermohaline features is quite limited in the Pacific.

Numerical simulations

The potential vorticity structure acts as a sensitive point of comparison between theory, numerical simulations and observations. Its relationship to the intensity of the flow, the mechanisms of injection of boundary conditions, and equilibration by eddy mixing, and its quasi-conservative 'tracer-like' nature make it an effective field to study, extending the basic dynamic topography maps.

Complementing the eddy-free, diffusive simulations with high vertical resolution and surface buoyancy forcing (see discussion of a/L_0 above) are the active eddy models with higher horizontal, but sparser vertical resolution²². These models contain explicitly the eddy transport of momentum that was invoked in the theory to predict homogenized gyres.

Here we show some results from a single numerical experiment based on quasigeostrophic equations in an eight-vertical-layer rectangular (1,400 km × 2,800 km) ocean basin. The ocean



©1984 Nature Publishing Group

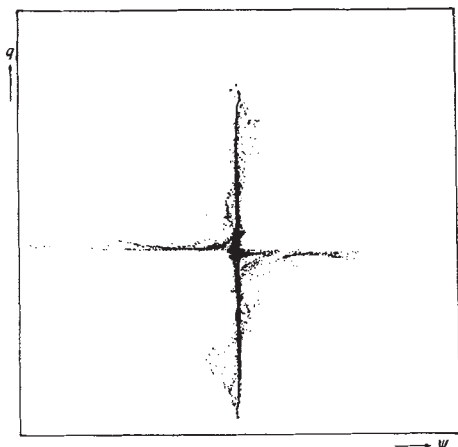


Fig. 8 Scatter plot of q against ψ for the mean fields on an interior layer (the third level of the current model). The pattern of a cross shows that either ψ vanishes (where the fluid is motionless) or ∇q vanishes (where the gyres are homogenized).

is driven at its surface by a symmetrical double-gyre wind stress, producing a western boundary current, its eastward extension at the middle latitude of the basin, and a Sverdrup regime in mid-ocean. There is no thermal forcing of the system and such a model does not allow for the outcrop of isopycnal surfaces; that is, all layers but the uppermost are 'buried' and feel the influence of the wind forcing only by way of eddy processes²². The eddy field arises due to a combined baroclinic/barotropic instability of the eastward jet, baroclinic instability of the tight westward recirculation that develops, and baroclinic instability of the distant westward return flow analogous, for example, to the North Equatorial Current. Very high horizontal resolution (~ 20 km) is needed to model these eddy generation processes adequately.

The experiment was carried out as an initial-value problem in which the ocean is spun up to a statistical equilibrium and the final state examined in terms of time-averaged mean fields (5-yr averaging time) and the deviations from them. The equations are just the quasi-geostrophic equivalent of the potential vorticity equation for an eight-layer ocean. The dissipation is modelled by a biharmonic lateral friction, $-A\nabla^6\psi$, in all layers and by a bottom friction $-\varepsilon\nabla^2\psi$, that exists in the lowest layer only. For boundary conditions on the side walls we take the streamfunction $\psi = \text{constant}$ (though a function of z and t), and $\nabla^2\psi = \nabla^4\psi = 0$.

Figure 7 shows maps at several levels in the model: the mean streamfunction $\bar{\psi}$ (Fig. 7a) and the mean potential vorticity $\bar{q}$ (Fig. 7b). The first, third and fifth of the eight levels are shown, corresponding to mean depths of 150, 850 and 1,750 m, respectively. The isopycnal depth of the middle level is shown in Fig. 7c, and the instantaneous potential vorticity of that level in Fig. 7d. Note first the nature of the mean circulation. In the upper layer the streamfunction shows the surface circulation to consist of twin gyres with a well-defined western boundary current, eastward extension, an inertial recirculation near the jet and a broad Sverdrup return flow. These are all features common to earlier calculations, but the two- and three-layer models were insufficiently resolved in the vertical to define the three-dimensional shape of the gyres.

At upper thermocline depths (Fig. 7a) the subtropical gyre has shrunk considerably poleward; the western boundary current and recirculation persist but are reduced in intensity. Finally, below the thermocline only the recirculation is present in the mean. The 200 cm s^{-1} eastward jet is far more baroclinic than the 25 cm s^{-1} westward recirculation, in accord with observations and the nature of the 'barotropization' process in stratified rotating fluids. Both the intensity of the circulation and the geometry of the free jet separating the major gyres are in general accord with observations.

Figure 7b shows the patterns of potential vorticity that are

produced by the dynamical and mixing processes occurring in the model. In the upper layer the isolines of $\bar{q}$ show both a very sharp frontal feature at the location of the eastward jet, as well as something of the gyre structure. The reversal of $\partial\bar{q}/\partial y$ in the distant westward return flow, coupled with the generally positive sense of $\partial\bar{q}/\partial y$ at deeper levels, suggest conditions favourable to baroclinic instability on the outer flanks of the gyres. Eddies are indeed generated in that region though the eddy energy there is an order of magnitude smaller than in the vicinity of the eastward jet.

In mid-depth layers (Fig. 7b) a vast region of homogenized potential vorticity exists. In this calculation the homogenized volume persists through several layers, representing a range of depths from just below the surface layer to the base of the thermocline (300–1,200 m). Below the thermocline the region of homogenization is reduced in size to coincide with the deep inertial recirculation.

The maps in Fig. 7b reveal that the sharp front in the upper layer is present in both instantaneous and mean potential vorticity. In the layers below (Fig. 7d) the instantaneous q is quite uniform (as expected from eddy-stirring arguments), yet we see streamers being swept into the region by the free jet. The tortuous deformations of the geostrophic contours are a familiar signature of the enstrophy cascade, and they emphasize the turbulent nature of even the outer reaches of the gyres. They are also evidence of the eddy momentum transport that drives the gyres to equilibrium. Figure 8, a plot of mean q and mean ψ for all the points in layer 3, shows the strong tendency for either $\bar{\psi}$ or $\nabla\bar{q}$ to vanish.

Finally, we consider the map of the depth of the middle density surface (Fig. 7c). Note the greatly different impression of the deep circulation given by the map of $\bar{h}$ and the maps of $\bar{\psi}$. The former is, of course, a measure of $\partial\bar{\psi}/\partial z$, the vertical shear of the circulation, and contains the same information as maps of dynamic topography. The gyres defined by $\bar{h}$ are much broader than the actual flow gyres, which contain a strong barotropic component. The parallel with our observational experience is clear: features like the westward inertial circulation are seen by current meters and floats more readily than in dynamic topography charts.

Although this simulation is highly idealized, it bears several points of resemblance to the field observations presented above. The vast pools of homogenized potential vorticity in the wind gyres strongly resemble those found in the Atlantic and Pacific. Near the surface there is a strong 'throughput' of potential vorticity due to forcing, in both model and oceans. A sharp front-like gradient of potential vorticity occurs at the gyre boundary in the Pacific (Fig. 5a) and in the model, (Fig. 7b) which expresses the impedance of the gyre boundary to eddy transport of q . At deeper levels in the Pacific, and in the model, the gyre boundary is invisible in the q distributions; with far weaker forcing and damping of q , eddy transport successfully mixes both gyres to the same value. The regions of reversal of the mean gradient of q , indicative of baroclinic instability, are similar in the model and in the oceanic observations (Fig. 7b; the relevant upper level maps are not shown here).

We feel that the three techniques of theory, analysis of observations, and numerical experiment described here have shown unusually close correspondence. A far closer synthesis of models and observations may now be possible than previously obtained. The kinematic and dynamic pictures of the oceans, as seen in trace chemistry and potential vorticity maps, respectively, complement one another strongly. Beyond the maps of the mean q field described here, it is clear that constructing maps of the instantaneous q field reveals much about the waves and geostrophic turbulence that dominate the transient energy spectrum of geophysical fluids^{23,24}. This technique, coupled with a lagrangian (particle-following) description, is of great potential value in dynamical studies of the oceans and atmosphere.

P.B.R. is supported by NSF grant OCE-83-19780, T.K. by the Center for Analysis of Marine Systems at WHOI, and W.R.H. by the NSF.

Received 2 September; accepted 4 December 1983.

1. Rhines, P. B. & Young, W. R. *J. mar. Res.* **40**, Suppl. 559–596, 849–872 (1982).
2. Welander, P. *J. mar. Res.* **29**, 60–68 (1971).
3. Hough, S. S. *Phil. Trans. R. Soc. A* **189**, 201–257 (1897).
4. Lamb, H. *Hydrodynamics* 6th edn (Dover, New York, 1932).
5. Montgomery, R. B. *Pap. phys. Oceanogr. Met.* **6**, (1938).
6. Reid, J. L. *Johns Hopkins oceanogr. Stud.* **2** (1965).
7. Behringer, D. W. thesis, Univ. California, San Diego (1972).
8. McDowell, S., Rhines, P. B. & Keffer, T. *J. phys. Oceanogr.* **12**, 1417–1436 (1982).
9. Sarmiento, J. L., Rooth, C. G. H. & Roether, W. *J. geophys. Res.* **87**, 8047–8056 (1982).
10. Rhines, P. B. & Young, W. R. *J. Fluid Mech.* **122**, 347–367 (1982).
11. Holland, W. R. & Rhines, P. B. *J. phys. Oceanogr.* **10**, 1010–1031 (1980).
12. Weiss, N. O. *Proc. R. Soc. A* **293**, 310–328 (1966).

13. Rhines, P. B. & Young, W. R. *J. Fluid Mech.* **133**, 135–148 (1983).
14. Dewar, W. K., Rhines, P. B. & Young, W. R. *Geophys. Astrophys. Fluid Dyn.* (in the press).
15. Luyten, J. R., Pedlosky, J. & Stommel, H. *J. phys. Oceanogr.* **13**, 292–309 (1983).
16. McCartney, M. S. *J. mar. Res.* **40**, Suppl. 427–464 (1982).
17. Cox, M. D. & Bryan, K. *J. phys. Oceanogr.* (in the press).
18. Levitus, S. *NOAA Prof. Pap.* **13**, (1983).
19. Sarmiento, J. L. *J. phys. Oceanogr.* (in the press).
20. Coats, D. A. *J. geophys. Res.* **86**, 8031–8036 (1981).
21. Reid, J. L. *J. geophys. Res.* **83**, 5063–5067 (1979).
22. Holland, W. R. *J. phys. Oceanogr.* **8**, 363–392 (1978).
23. Rhines, P. B. in *The Sea* Vol. 6 (ed. Goldberg, E. N.) 189–319 (Wiley-Interscience, New York, 1977).
24. McIntyre, M. E. & Palmer, T. N. *Nature* **305**, 593–600 (1983).

Single base deletion in the vasopressin gene is the cause of diabetes insipidus in Brattleboro rats

Hartwig Schmale & Dietmar Richter

Abteilung Zellbiochemie, Institut für Physiologische Chemie, Universität Hamburg, Martinistrasse 52, D-2000 Hamburg 20, FRG

In rats with hereditary hypothalamic diabetes insipidus (Brattleboro rats) the gene for the vasopressin precursor lacks a single G residue in the protein-coding region. The mutation gives rise to an open reading frame predicting a hormone precursor having a different C-terminus.

STUDIES of hereditary hypothalamic diabetes insipidus in mutant (Brattleboro) rats show that these animals lack the hormone vasopressin as well as its corresponding neurophysin carrier^{1,2}. Hence, resorption of water in the distal kidney tubuli, which is controlled by vasopressin, is greatly impaired. Therefore, animals with this disease show a high water uptake accompanied by excretion of an excessive amount of dilute urine. These symptoms can be reversed by administering arginine vasopressin (AVP) subcutaneously³.

In normal rats the vasopressin gene is transcribed in magnocellular neurones in both the supraoptic and paraventricular nuclei of the hypothalamus⁴; the synthesized precursor is packaged into neurosecretory granules and transported axonally in the stalk to the posterior pituitary^{4,5}. It is now clear that the precursor consists of the hormone, its neurophysin carrier and a glycoprotein of as yet unknown function^{6,7}.

In Brattleboro rats, neurones of the supraoptic and paraventricular nuclei are hypertrophied⁸ and those cells believed to produce vasopressin lack normal neurosecretory granules⁹; instead, these cells produce smaller, dense-cored granules which suggests that, nevertheless, the machinery for packaging secretory material is intact. As small granules have been observed also in normal rats it appears unlikely that they contain a mutated vasopressin precursor. A candidate for this role, however, could be the 'peptide X' isolated in small quantities from Brattleboro rats and which has neurophysin- and vasopressin-like immunoreactivity⁴. This peptide is not glycosylated.

Production of oxytocin, the other neurohypophyseal hormone which is synthesized together with a second neurophysin carrier as a composite precursor^{10,11}, seems to be unaffected by the Brattleboro mutation. The observed lower levels of oxytocin in the neurohypophysis of Brattleboro rats are probably due to osmotic stress, as normal levels are restored when the animals are treated with vasopressin¹².

To elucidate the molecular basis of the genetic defect in vasopressin production, we studied rats homozygous for diabetes insipidus. Based on recent studies, the gene for the vasopressin precursor in normal rats is interrupted by two intervening sequences which separate three distinct exons encoding the three functional domains of the precursor—vasopressin, neurophysin and glycoprotein⁷. A similar arrangement has been observed

for the oxytocin gene except that a sequence encoding a glycoprotein is absent¹¹. It is possible to derive from these genes DNA probes which are specific for vasopressin or oxytocin and hence can be used to discriminate between the two genes.

We report here the identification and sequence analysis of the vasopressin gene from rats with diabetes insipidus.

Isolation and sequence analysis

To identify the mutant gene, liver DNA from Brattleboro rats was cleaved with different restriction enzymes, blotted according to Southern¹³ and hybridized to a ³²P-labelled DNA fragment encoding the vasopressin precursor from normal rats. Our earlier restriction analysis of the normal rat gene with restriction enzymes *EcoRI* and *HindIII* revealed bands of 8.2 kilobases (kb) and 3.8 kb, respectively, both containing the complete vasopressin gene⁷. The Southern hybridization patterns of DNA fragments from normal and Brattleboro rats are identical (Fig. 1).

For further analysis, the 8.2-kb *EcoRI* DNA fragment from Brattleboro rats was cloned in the λ phage 641. Plaques (2×10^5) were screened with the DNA probe encoding the vasopressin precursor. Five plaques gave a positive hybridization signal. The *HindIII* DNA fragment of one of the identified clones was subcloned in the plasmid pUC8, mapped with restriction enzymes and sequenced according to the method of Maxam and Gilbert¹⁴. Figures 2 and 3 show that the sequences of the vasopressin genes from normal and Brattleboro rats are identical except for a deletion of a G residue in the second exon encoding the highly conserved part of neurophysin. This deletion causes a shift in the reading frame, predicting a vasopressin precursor having an entirely different C-terminus.

To exclude possible cloning artefacts, another of the five identified λ phage clones, in which the inserted DNA was in the opposite orientation, was analysed without prior subcloning. Part of the *HindIII* DNA fragment of this clone containing the site of deletion was directly sequenced and gave the same result as in Fig. 2.

Brattleboro mRNA

To determine whether the mutant vasopressin gene is transcribed and the primary transcript correctly spliced,

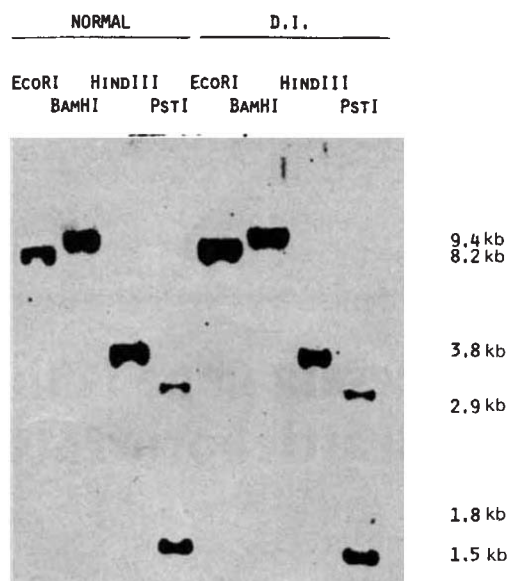


Fig. 1 Hybridization of a ³²P-labelled vasopressin DNA probe to genomic DNA, digested with different restriction endonucleases, from normal and diabetes insipidus (D.I.) rats.

Methods: The homozygosity of Brattleboro rats (obtained from Holland Centraal Proefdierenbedrijf, Zeist, The Netherlands) was checked by measuring water uptake and excretion over several days. High-molecular weight DNA was prepared from the rat liver²¹, then 10 µg were digested with the indicated restriction endonucleases, electrophoresed on a 0.8% agarose gel, denatured and transferred to nitrocellulose sheets essentially as described by Southern¹³. Hybridization was performed at 65 °C with the ³²P-labelled subclone pVλHindIII comprising the entire vasopressin gene from normal rats⁷. In the hybridization conditions used only the vasopressin, not the oxytocin, gene gave a positive signal. The nitrocellulose blots were washed at 65 °C with 0.1 × SSC, 0.1% SDS and exposed to X-ray film for 3 days.

hypothalamic poly(A)⁺ RNA was separated on agarose gels in denaturing conditions, blotted onto a nitrocellulose filter and hybridized to a vasopressin-specific DNA probe⁷. As shown in Fig. 4a, the vasopressin mRNAs from both normal and Brattleboro rats migrate with a size of ~700 nucleotides. This strongly suggests that the primary transcripts are similarly spliced in normal and Brattleboro rats. Normal and diabetes insipidus rats contain comparable amounts of vasopressin and oxytocin precursor-specific mRNA. The sizes of AVP mRNA (700 bases and oxytocin mRNA (650 bases) are not altered in the case of the mutant rat (Fig. 4a, c).

The start site of transcription of the normal vasopressin gene has been determined by S₁ nuclease mapping⁷. Besides the start site beginning with an A residue, a second protected fragment was observed, indicating microheterogeneity of the mRNA. S₁ mapping of the mRNA from Brattleboro rats gave identical results, excluding the possibility of alternative transcription start sites which could affect the secondary structure of the mRNA and thus its translation efficiency.

To show that the single base deletion determined from the sequence studies of genomic DNA was also present in the mRNA encoding the predicted precursor, poly(A)⁺ RNA from Brattleboro rats was analysed by an oligonucleotide hybridization assay, which allows the detection of a single mismatched base pair¹⁵. An 18mer oligonucleotide (5'-CAGCGGCCTCGCTCCGC-3') was synthesized that is complementary to the altered sequence in the Brattleboro rat, with nine nucleotides both upstream and downstream of the deletion site, thus differing from the normal rat sequence by one nucleotide. According to Wallace *et al.*¹⁵, differences in a single nucleotide can be demonstrated by applying specific hybridization conditions. The Northern blot analysis of poly(A)⁺ RNA hybridized to the ³²P-labelled

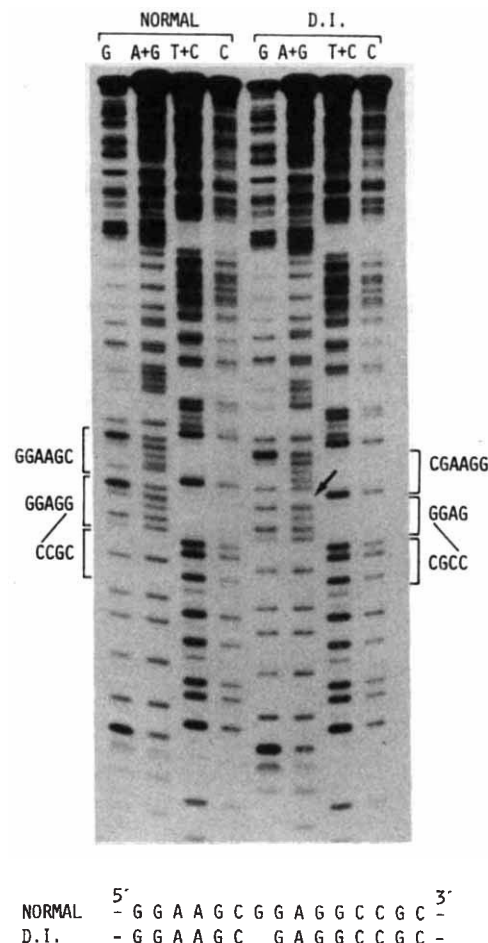


Fig. 2 Gene sequence analysis indicating the site of deletion in the diabetes insipidus rat. The DNA sequence is given below the autoradiogram of the gel, showing the deletion of a single G residue (arrowed in the autoradiogram). D.I., diabetes insipidus rats.

Methods: The 8.2-kb *EcoRI* DNA fragment from Brattleboro rats (Fig. 1) was cloned into the insertion vector λ641 (ref. 22). 35 µg of liver DNA were digested with *EcoRI* and run on a 0.8% low melting point agarose gel. DNA fragments of 7–9 kb (according to the ethidium bromide-stained marker DNA) were extracted with phenol/chloroform and ligated into the single *EcoRI* site of λ641 which can accept inserts of up to ~11 kb²². The recombinant DNA was packaged *in vitro* using a kit (Amersham). When using *Escherichia coli* POP 13b as the indicator strain, only recombinant phages yield visible plaques. After transfer of 2 × 10⁵ plaques to nitrocellulose sheets, hybridization with a ³²P-labelled vasopressin DNA probe⁷ yielded five clones containing the 8.2-kb fragment. Two types of clones were identified by restriction mapping as containing the insert in both orientations. The 3.8-kb *HindIII* fragment was cut out and subcloned in the plasmid pUC8. A similar subclone was obtained from a genomic library of normal rats⁷. For sequence comparison between the normal and Brattleboro rat gene, the sequencing reactions²¹ and the sequencing gels were run in parallel.

18mer oligonucleotide shows that only the mRNA from Brattleboro rats gave a positive hybridization signal (Fig. 4b). This experiment independently confirms at the mRNA level the single deletion identified in the Brattleboro rat gene sequence.

In vitro studies

As the Northern blot analysis indicated a similar sized mRNA from both normal and mutant rats, attempts were made to study expression of this gene in a heterologous system. Microinjection of the vasopressin gene from normal and Brattleboro rats into the nucleus of frog oocytes and dot-blot analysis indicated that the genes are transcribed, however translation products cross-

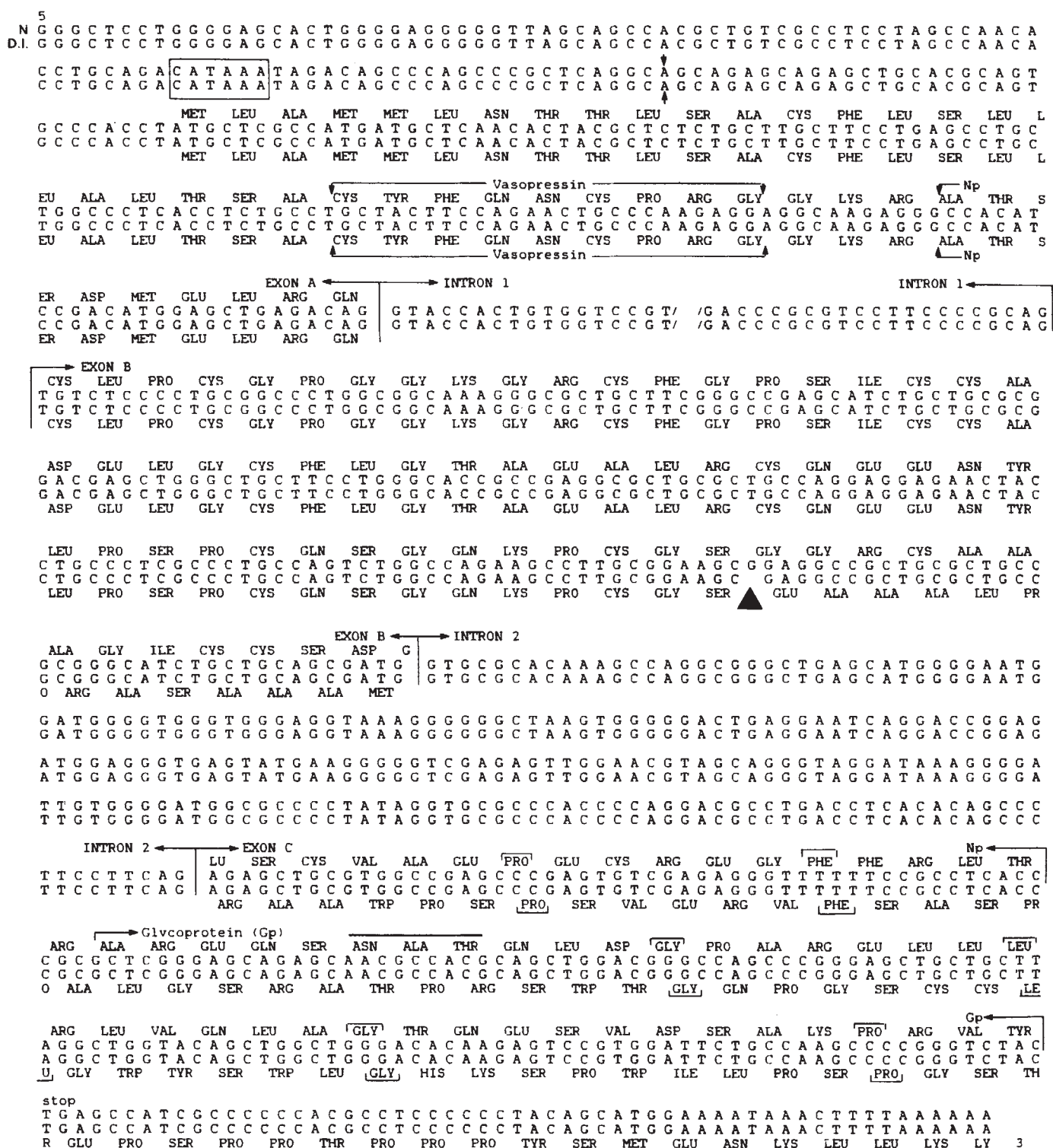


Fig. 3 Comparison of the gene sequences from normal (N) and diabetes insipidus (D.I.) rats. The nucleotide sequence of the rat anti-sense strand is shown together with the deduced amino acid sequence. The Goldberg-Hogness sequence²³ is boxed. The vertical arrowheads indicate the start site of transcription determined by S₁ nuclease mapping. The triangle indicates the position of the G residue deletion. Intron 1 is only partially shown. The thick line above the amino acid sequence of exon C in the normal gene shows the glycosylation site. Amino acid positions that are identical in the C-termini of the deduced precursors are marked by a frame (—). Np, neurophysin; Gp, glycoprotein.

reacting with antibodies raised against vasopressin have not yet been detected.

When hypothalamic mRNA from Brattleboro rats was translated in a cell-free system from reticulocyte lysate, minute amounts of a vasopressin precursor (~1% of that in normal rats) could be identified with anti-vasopressin antiserum (data

not shown). The low yield of synthesized material is probably due to the lack of a stop codon. As indicated above, the sequence analysis predicts an open reading frame up to and beyond the polyadenylation site. Theoretically, translation of the mRNA could lead to a product having a polylysine tail at the C-terminus of this precursor. On the other hand, lack of the stop codon

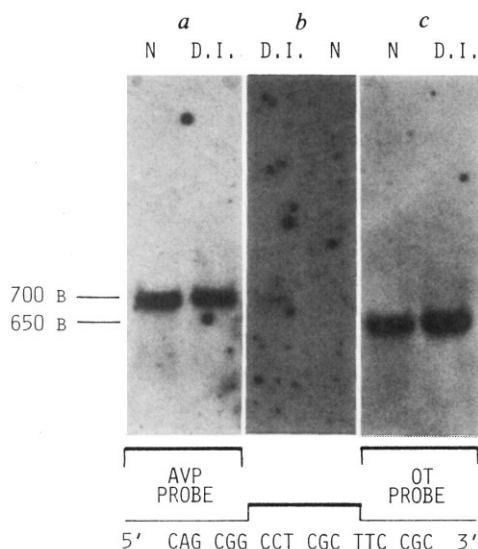


Fig. 4 Northern blot analysis of poly(A)⁺ RNA from normal (N) and diabetes insipidus (D.I.) rats. Hypothalami from normal and Brattleboro rats were homogenized in guanidinium thiocyanate and the RNA purified by centrifugation through a CsCl cushion²⁰. Polyadenylated RNA was prepared by oligo(dT)-cellulose chromatography. Poly(A)⁺ RNA (5 µg) was separated in denaturing conditions on a vertical 1.2% agarose gel containing 10 mM methylmercuric hydroxide and transferred to a nitrocellulose filter using 3 M NaCl, 0.3 M Na-citrate (20×SSC). The blot to be hybridized sequentially with the AVP or oxytocin (OT) probe was prehybridized overnight in 50% formamide, 50 mM Na-phosphate pH 7.0, 0.9 M NaCl, 5 mM EDTA, 0.1% SDS, 1×Denhardt's and 500 µg ml⁻¹ denatured herring sperm DNA at 42 °C (ref. 24). The first hybridization (a) was carried out in the same conditions using 10⁶ c.p.m. ml⁻¹ of the ³²P-labelled 670 base pair-long insert of the AVP-specific subclone pVAPstI⁷. The nitrocellulose filter was washed for 20 min at 42 °C and then for 20 min at 50 °C using the prehybridization solution without Denhardt's and herring sperm DNA and finally for 20 min in 0.1% SSC, 0.1% SDS at 50 °C. After dehybridization by washing the nitrocellulose filter in 3×SET, 50% formamide, 1×Denhardt's, 0.1% SDS at 68 °C for 1 h, a second hybridization was done in the conditions described but now using the ³²P-labelled 650 base pair-long *Ava*I-*Eco*RI restriction fragment (c) specific for the rat oxytocin gene¹¹. For the oligonucleotide hybridization (b) the 18mer 5'-CAGCGGCCTCGCTTCCGC-3' (Applied Biosystems) complementary to the Brattleboro DNA sequence around the deleted G residue (Fig. 2) was labelled with [γ -³²P]ATP by T4 kinase (specific activity 6.3×10⁸ c.p.m. µg⁻¹; Boehringer) and separated from the other reaction products as described by Kidd *et al.*²⁵. After prehybridization in 4×SSC, 0.1% SDS, 1×Denhardt's, 10 µg ml⁻¹ denatured herring sperm DNA, 10 mM EDTA at 35 °C for 1 h, the Northern blot was hybridized with the ³²P-labelled 18mer oligonucleotides in the same solution (10⁶ c.p.m. ml⁻¹) at 35 °C for 48 h. The blot was washed in four changes of 5×SSC, 0.1% SDS for 5 min each at 4 °C.

could prevent normal run-off of the ribosomes, blocking them on the mRNA.

Discussion

We have described the molecular basis for the hereditary defect in the vasopressin gene of rats with diabetes insipidus. The defect is due to a single deletion in the sequence encoding the neurophysin and gives rise to an open reading frame.

This new reading frame predicts a greatly altered vasopressin precursor (Fig. 5): not only is the glycosylation site absent but the C-terminus of the neurophysin is drastically changed. From amino acid residue 64 of the neurophysin up to position 95, all amino acid residues have been changed except for two. These changes are particularly significant from residue 64 to 76, encoded by exon B⁷, a region normally belonging to the highly

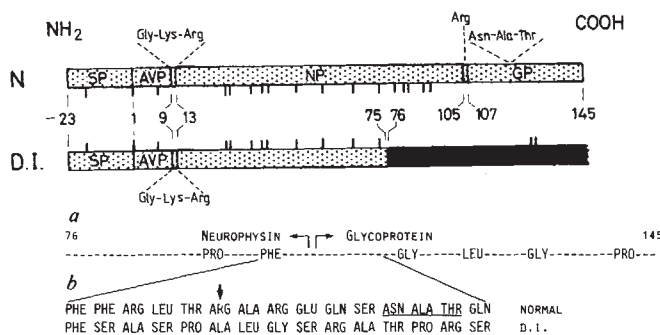


Fig. 5 Comparison of the vasopressin precursors from normal (N) and diabetes insipidus (D.I.) rats deduced from gene structures. The negative number indicates the translation start site; the other numbers show significant positions within the prohormone: 1, first amino acid (cysteine residue) of the prohormone; 9–13, the amino acids Gly, Lys, Arg separating the hormone from the neurophysin; 75–76, the start site of the new reading frame in the predicted D.I. precursor; 105, last amino acid residue of the neurophysin; 107, first amino acid residue of the glycoprotein; 145, last amino acid of the prohormone. The basic amino acid (arginine) separating neurophysin and the glycoprotein, and the glycosylation site (Asn-Ala-Thr) are indicated. The vertical lines beneath (N) or on top of (D.I.) the precursor schemes show the positions of the cysteine residues. SP, signal peptide; AVP, arginine vasopressin; NP, neurophysin; GP, glycoprotein. **a**, Partial amino acid sequence of the vasopressin precursor from normal rats from residue 76 to residue 145 of the prohormone. Only those amino acids which occur at identical positions in both precursors are listed; dashes indicate residues which are different. **b**, Sequence comparison of the region in which the neurophysin adjoins the glycoprotein including the glycosylation site (underlined). The latter is absent in the D.I. precursor; possible alternative processing sites are indicated by the two small arrows.

conserved part of the neurophysin. Of 14 cysteine residues present originally in the neurophysin, five have been replaced by other amino acids (Fig. 5)—this should have a severe effect on the folding of the mutant neurophysin¹⁶.

The new reading frame of the mutant gene also predicts the replacement of a basic amino acid (arginine) normally separating the neurophysin from the glycoprotein and which is known to direct proteolytic cleavage. An alternative cleavage site could occur at position 110 of the mutant precursor (Fig. 3, 5); this would lead to a neurophysin-like protein extended for four amino acids.

The failure to detect vasopressin and its associated neurophysin^{1,8} and glycoprotein¹⁷ in Brattleboro rats by radio-immunoassays or by immunohistochemical studies could be explained by postulating that the mRNA is not translated or that any translation product either remains unprocessed or is rapidly degraded. As the most significant change in the predicted precursor sequence is the lack of a stop codon, protein synthesis may come to a halt. It may be that the vasopressin mRNA is translated on the membranes of the rough endoplasmic reticulum, but because of the missing stop codon, the ribosomes cannot leave the mRNA. Whether such a halted ribosome-mRNA complex with a nascent, yet incomplete peptide chain may interfere with the synthesis of other secretory proteins can only be speculated. In this context it is interesting that the vasopressin-producing cells which also synthesize Leu-enkephalin contain significantly lower amounts of this peptide in Brattleboro rats¹⁸.

Alternatively, the predicted precursor may be synthesized but in very small quantities. It is possible that 'peptide X'¹⁹ isolated from Brattleboro hypothalami is identical with the predicted precursor, as it appears to have antigenic sites for vasopressin and neurophysin and lacks a carbohydrate chain. If the predicted precursor is in fact synthesized, lack of the glycoprotein moiety as well as the drastically altered C-terminus of the

neurophysin may hinder its further packaging and processing in the neurosecretory granules. On the other hand, the modified C-terminus may have caused the neurophysin to have lost one of its supposed functions, namely to protect the hormone from proteolytic degradation²⁰.

We thank Heidje Christiansen for technical assistance and Drs Monika Rehbein for DNA preparation, Gerd Scherer for providing λ 641 and Richard Ivell for helpful discussions. This work received financial support from the Deutsche Forschungsgemeinschaft.

Received 16 January; accepted 25 February 1984.

1. Valtin, H., Stewart, J. & Sokol, H. W. *Handbk Physiol.* **7**, 131-171 (1974).
2. Pickering, B. T. & North, W. G. *Ann. N. Y. Acad. Sci.* **394**, 72-81 (1982).
3. Cheng, S. W. T., North, W. G. & Gellai, M. *Ann. N. Y. Acad. Sci.* **394**, 437-480 (1982).
4. Brownstein, J. M., Russell, J. T. & Gainer, H. *Science* **207**, 373-378 (1980).
5. Gainer, H. *Prog. Brain Res.* **60**, 205-215 (1982).
6. Land, H., Schütz, G., Schmale, H. & Richter, D. *Nature* **295**, 299-303 (1982).
7. Schmale, H., Heinsohn, S. & Richter, D. *EMBO J.* **2**, 763-767 (1983).
8. Sokol, H. W. & Valtin, H. *Ann. N. Y. Acad. Sci.* **394**, 1-828 (1982).
9. Valtin, H., North, W. G., LaRochelle, F. T., Sokol, H. W. & Morris, J. *Proc. 7th Int. Congr. Nephrol.*, Basel, 313-320 (Karger, Basel, 1978).
10. Land, H. *et al. Nature* **302**, 342-344 (1983).
11. Ivell, R. & Richter, D. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
12. Valtin, H. & Schroeder, H. A. *Am. J. Physiol.* **206**, 425-430 (1964).
13. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
14. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
15. Wallace, R. B. *et al. Nucleic Acids Res.* **6**, 3543-3557 (1979).
16. Breslow, E. A. *Rev. Biochem.* **48**, 251-274 (1979).
17. Lu, C.-L., Cantin, M., Seidah, N. G. & Chrétien, M. *J. Histochem. Cytochem.* **30**, 999-1003 (1982).
18. Martin, R. & Voigt, K. H. *Nature* **289**, 502-504 (1981).
19. Russell, J. T., Brownstein, M. J. & Gainer, H. *Endocrinology* **107**, 1880-1881 (1980).
20. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
21. Blin, N. & Stafford, D. W. *Nucleic Acids Res.* **3**, 2303-2308 (1976).
22. Murray, N. E., Brammar, W. J. & Murray, K. *Molec. gen. Genet.* **150**, 53-61 (1977).
23. Breathnach, R. & Chambon, P. A. *Rev. Biochem.* **50**, 349-383 (1981).
24. Gal, A., Nahon, J.-L. & Sala-Trepat, J. M. *Analyt. Biochem.* **132**, 190-194 (1983).
25. Kidd, V. J., Wallace, R. B., Itakura, K. & Woo, S.L.C. *Nature* **304**, 230-234 (1983).

LETTERS TO NATURE

Terrestrial mass extinctions, cometary impacts and the Sun's motion perpendicular to the galactic plane

Michael R. Rampino* & Richard B. Stothers

National Aeronautics and Space Administration,
Goddard Institute for Space Studies, 2880 Broadway, New York,
New York 10025, USA

Episodes of mass extinctions on the Earth are now strongly suspected to be cyclical¹. We report here that our analysis of the data of Raup and Sepkoski¹ suggests that the dominant cyclicity in major marine mass extinctions during at least the past 250 Myr is 30 ± 1 Myr, with the standard deviation of an individual episode being ± 9 Myr. We find this terrestrial cycle to be strongly correlated with the time needed for the Solar System to oscillate vertically about the plane of the Galaxy, which is 33 ± 3 Myr according to the best current astronomical evidence. It is argued that galactic triggering or forcing of terrestrial biological crises may arise as a result of collisions (or close encounters) of the Solar System with intermediate-sized to large-sized interstellar clouds of gas and dust, which are sufficiently concentrated towards the galactic plane to produce the observed cyclicity and its scatter. Among other consequences, a nearby interstellar cloud would gravitationally perturb the Solar System's family of comets and thereby increase the flux of comets and comet-derived bodies near the Earth, leading to large-body impacts. We find a dominant cyclicity of 31 ± 1 Myr in the observed age distribution of impact craters on Earth, the phase of this cycle agreeing with that shown by the major biological crises. Our galactic hypothesis can thus simultaneously account for the mean interval between major terrestrial crises and for the 50% scatter of the time intervals about their mean value.

Raup and Sepkoski¹ have very recently presented evidence for an approximate cyclicity in marine mass extinctions over the past 250 Myr. Fourier analysis of their data showed a dominant periodicity of 30 Myr, a best-fit curve yielded a cycle of 26 Myr, and a non-parametric test (previously developed independently and somewhat differently in ref. 2) revealed a significant cycle at 26 Myr and an only slightly less significant cycle at 30 Myr. Using less extensive data, Fischer and Arthur³ had already suggested a 32-Myr periodicity in marine mass extinctions.

Which periodicity, 26 Myr or 30 Myr, should be preferred?

Raup and Sepkoski listed all discernible mass extinction peaks that exceeded 2% extinctions on a family level. As they were well aware, however, there are a number of uncertainties in their selection procedure. First, 2% may be too small to be statistically meaningful; Raup⁴ elsewhere suggested the use of 10%. Second, there are several minor fluctuations on the time curve of mass extinctions, some of which Raup and Sepkoski selected as true peaks and others they rejected as spurious peaks. Third, as the resolution in time is only one geological stage, other individual minor mass extinction episodes may exist at the substage level. In view of these uncertainties concerning identification and significance, we consider it best to follow Raup⁴ and adopt a cutoff of 10% extinctions, while disregarding all the minor inflections on the curve. The large remaining peaks then refer exclusively to the major mass extinction episodes and can be assumed to be completely sampled. They are listed in Table 1 (where the dates have been slightly revised according to ref. 5). Nine dates appear in this table, in contrast to the 12 listed by Raup and Sepkoski.

Intervals of time between successive episodes of major mass extinctions lie in the range 17-53 Myr, with mean time interval 29 Myr. We regard the large (50%) dispersion as physically real, although some of it must be due to uncertainties of the dating. Reasons given below suggest that the dispersion is randomly generated. If, however, a true underlying mean periodicity does exist and no major mass extinction episodes are either missing or redundant, we can unambiguously assign a cycle number to each episode (Table 1). Non-parametric tests (like those in refs 1, 2) are then no longer appropriate. Regression of the episode date on the cycle number by the method of least squares provides an unbiased estimate of the best-fitting mean period, which is found to be 30 ± 1 Myr. If other recently proposed geological time scales are adopted¹, essentially the same mean period emerges. This invariance is not surprising because the central limit theorem ensures that the presence of even rather large random errors in the dates will not significantly affect either the mean time interval or the best-fitting least-squares period.

If the cycle number is left as a free parameter, the best-fitting mean period P derived from a suitable non-parametric method², in which the observed times are fitted to a formula of the type $t = t_0 + nP$, is 26 Myr, where t , t_0 and n are an observed time, the most recent epoch and an integer, respectively. This would agree with Raup and Sepkoski's¹ result. However, Raup and Sepkoski have emphasized the approximate nature of any period that can be derived from so few data points. We prefer, in fact, the assignment of the cycle numbers as given above and therefore the period of 30 Myr. There are additional reasons for our preference: first, other kinds of geological data (including 41 dated impact craters) show periods of 30-35 Myr, and, second,

* Present address: Department of Geological Sciences, Columbia University, New York, New York 10027, USA.

the larger variance produced with the longer period actually appears to have a natural physical basis, as will be discussed below.

A cyclicity of the specific length and rough regularity shown by the marine mass extinction data is difficult to reconcile with a purely terrestrial triggering mechanism^{1,6}. There is also no known solar mechanism that could operate on this long a time scale⁷. A galactic mechanism therefore appears to be the most likely. A number of cycles ranging from 100 to 900 Myr (refs 8–13), are associated with the slow revolution of the Solar System in the plane of the Galaxy about the galactic centre. These cycles, however, are much too long to explain the 30-Myr spacing of the major extinction episodes.

Several authors^{13–16} have pointed out that the Solar System also executes a nearly periodic motion perpendicular to the galactic plane (in the z direction). Innanen *et al.*¹³ have calculated the Sun's orbit backward in time in a simple, but physically reasonable, model of the Galaxy, and have found the average period for one complete vertical oscillation to be ~ 67 Myr. They remarked (as had Hatfield and Camp¹⁵) that a period of this length is similar to the known lengths of the geological periods. The physically meaningful cycles, however, are the dynamical half-periods, $P_{1/2} \approx 33$ Myr (plane crossing to plane crossing). The calculated crossing times of Innanen *et al.* are given in Table 1.

For a slowly moving object like the Sun, the vertical oscillation is essentially a simple harmonic motion governed by the local galactic mass density ρ_0 . Innanen *et al.* adopted $\rho_0 = 0.20 M_\odot \text{pc}^{-3}$, which agrees with recent values of $\rho_0 = 0.14\text{--}0.21 M_\odot \text{pc}^{-3}$ determined from dynamical studies of nearby stars^{17–21}. By using the harmonic relation $P_{1/2}\rho_0^{1/2} = \text{constant}$, we estimate the possible error associated with $P_{1/2} = 33$ Myr as less than ± 3 Myr.

The galactic time series in Table 1 can now be compared more rigorously with the terrestrial mass extinction time series. The correlation coefficient computed for the two time series in columns 2 and 3 of Table 1 is very high ($r = 0.996$). Student's t -test for matched pairs²² (successive time intervals matched between columns 2 and 3) confirms that the mean time interval in the terrestrial series does not differ significantly from the mean time interval, or half-period, of galactoververtical oscillation (the probability that $t = 0.91$ with 7 d.f. is $p = 0.40$). No mean phase difference between the two series is detectable, but some individual phase differences are large. Both the tight mean correlation and the significant individual phase differences suggest the following astrophysical model.

Passage of the Sun through the galactic disk involves the Solar System in varying amounts of diffuse interstellar material whose density falls off roughly exponentially above and below the galactic plane,

$$N(z) = N_0 \exp(-|z|/\beta) \quad (1)$$

But the Sun's total vertical trajectory is only ~ 70 pc (ref. 13), which is less than the vertical scale height $\beta \approx 120$ pc for ordinary interstellar gas and dust clouds²³ ($n = 10^1\text{--}10^2$ particles cm^{-3})²⁴ and less than $\beta \approx 100$ pc for the clouds ejected by supernova explosions^{25,26}. Therefore, the Solar System has an almost constant probability of hitting one of these clouds. Because of this, no real periodicity between encounters with either ordinary or supernova-produced clouds could exist. Moreover, significant collisions with supernova ejecta would probably occur less frequently than once every 50 Myr (refs 27–30). Giant molecular clouds (with $n > 10^3$ particles cm^{-3}) are more closely concentrated towards the galactic plane ($\beta \approx 15$ pc assuming $\beta = 2^{-1/2}\langle z^2 \rangle^{1/2}$)³¹, but their encounters with the Sun are expected to occur less frequently than once every few hundred Myr (refs 6, 32, 33).

Unless the giant molecular clouds⁶ or other very dense interstellar clouds²⁴ are much more common than presently thought, the most likely perturbing objects are intermediate-sized clouds. These have $n = 10^2\text{--}10^3$ particles cm^{-3} , radius $R = 3\text{--}6$ pc and mass $M = 10^3\text{--}10^4 M_\odot$ (ref. 24). They seem to be adequately concentrated to the galactic plane, with $\beta \approx 44$ pc (ref. 31). They

Table 1 Dates (Myr BP) of terrestrial and galactic events

Geological age ¹	Mass extinctions ¹	Galactic plane crossings ¹³	Difference	Cycle no.
Middle Miocene	11*	~ 0	+11	0
Late Eocene	37	31	+6	1
Maastrichtian	66	64	+2	2
Cenomanian	91	100	−9	3
Tithonian	144	135	+9	4
Bajocian	176	166	+10	5
Pliensbachian	193	197	−4	6
Norian	217	227	−10	7
Dzhulfian	245	259	−14	8

The mass extinctions listed here are the major events discussed in the text. Mass extinction dates have been revised after ref. 5. Columns 2 and 3 are not expected to agree in detail, for reasons given in the text.

* The magnitude of this episode is uncertain.

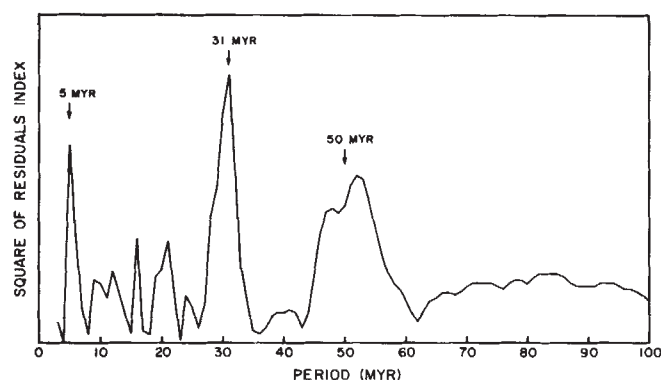


Fig. 1 Terrestrial impact craters: the square of the residuals index (defined in ref. 2) as a measure of the goodness of fit for trial periods that have been fitted to the observed age distribution of 41 craters over the interval $t = 1\text{--}250$ Myr (ref. 49).

also appear to be more or less uniformly distributed along the galactic plane³⁴; at least they do not show the strong concentration to spiral arms that is characteristic of the largest clouds³⁵.

To obtain a direct comparison with the vertical distribution of the clouds, the amount of time spent by the Sun at different intervals of height above or below the galactic plane is shown in Table 2; this has been computed by using an analytical integral of the galactoververtical orbit equation³⁶

$$t(z) = (P_{1/2}/\pi) \sin^{-1}(z/z_{\text{max}}) \quad (2)$$

One-third of the Sun's time is spent at $|z| > 60$ pc. This almost guarantees an approximate cyclicity in the long-term rate of collisions with intermediate-sized and large-sized clouds, which will occur preferentially at lower $|z|$ distances, although not always in the galactic plane itself. Using the cloud statistics of Talbot and Newman²⁴, we estimate about 1 collision per solar passage through the galactic disk. If only tidal (near) encounters with clouds have to be considered, the rate will, of course, be higher.

The one-third of the Sun's time that is spent farthest from the galactic plane is expected to be associated with only 20% of all the collisions and tidal encounters with clouds whose $\beta = 44$ pc, but 33% if $\beta > 70$ pc. From the data in Table 1, we find that 11% (1 out of 9 as compared with the expected 2 out of 9) among the major terrestrial extinction episodes actually occur at this unfavoured time. Within the statistical uncertainty, the agreement with our model is good.

Although the Sun at present lies very close to the galactic plane (8 ± 12 pc north), it is passing through a locally poor region of gas and dust^{23,27}. It is expected to move out of the region after ~ 3 Myr, at which time it will again become vulnerable to encounters with the intermediate-sized and large-sized interstellar clouds.

Table 2 Relative amounts of time spent by the Sun in successive intervals of galactic height

$ z $ (pc)	$z/z_{\max}$	$\Delta t/P_{1/2}$
0-7	0-0.1	0.06
7-14	0.1-0.2	0.06
14-21	0.2-0.3	0.07
21-28	0.3-0.4	0.07
28-35	0.4-0.5	0.07
35-42	0.5-0.6	0.08
42-49	0.6-0.7	0.08
49-56	0.7-0.8	0.10
56-63	0.8-0.9	0.12
63-70	0.9-1.0	0.29

Two consequences of collisions and tidal encounters with such clouds that may have large ultimate biological impact have been suggested. First, if actual penetration of the cloud occurs, the cloud particle density of $n > 10^2 \text{ cm}^{-3}$ would probably shut off the solar wind at the Earth's distance, as Begelman and Rees³⁸ have shown. Such a cloud could then directly pollute the Earth's upper atmosphere with hydrogen gas, leading to a variety of possible climatic effects³⁹⁻⁴¹. A larger cloud density of $n > 10^3 \text{ cm}^{-3}$ would also raise the Sun's luminosity significantly through gravitational accretion and so would directly affect insolation on the Earth^{6,12}. To traverse a cloud of radius 5 pc at a relative speed of 20 km s^{-1} requires 0.5 Myr.

Second, and probably more importantly, even if an actual collision with a cloud does not occur, the cloud's huge mass would gravitationally perturb the Solar System's inner reservoir of comets⁴² (as well as the surrounding Oort⁴³ comet cloud) and so focus a shower of comets on the innermost regions of the Solar System. According to equation (11) of Hills⁴², comets having semi-major axes equal to the outer radius ($\sim 2 \times 10^4 \text{ AU}$) of the inner cometary reservoir will burst into the vicinity of the Earth's orbit if an interstellar cloud of mass 10^3 – $10^4 M_{\odot}$ approaches within 5–15 pc of the Sun. (Smaller interstellar clouds, because of the diffuseness of their mass, could effectively perturb only the Oort comet cloud and therefore would focus only a few comets to small perihelion distances, while significant encounters of the Sun with passing stars would occur only once every 500 Myr.) According to data from Hills, during the calculated 10 Myr lifetime of the expected comet shower⁴², probably one large comet and several smaller comets would hit the Earth. Collision of a comet or a comet-derived asteroidal body with the Earth could have severe biological (and other geological) consequences⁴⁴⁻⁴⁸.

Grieve⁴⁹ has recently compiled the ages of all dated Phanerozoic (younger than 600 Myr) impact craters. By applying the above mentioned non-parametric method of time-series analysis² to the data in his Table 1 and text and excluding ages given only as upper limits, we find strong signals at periods of 31, 5 and ~ 50 Myr, as shown in Fig. 1 for the time interval $t = 1$ –250 Myr. These periods are significant at the 1% level, as we have determined by generating and analysing with the same method 5,000 random time series containing the same number of dates. The two periods of 5 and ~ 50 Myr are artefacts produced by the rounded numbers given for the ages of (mostly older) craters; 61% of the 41 crater ages ($t = 1$ –250 Myr) are divisible by 5, while 22% are divisible by 50. The spectral peak around 31 Myr, however, is consistently high in all of our time-series tests, which include the ranges $t = 1$ –250 Myr (41 craters), $t = 30$ –250 Myr (29 craters), $t = 1$ –600 Myr (62 craters) and $t = 250$ –600 Myr (22 craters).

Some of the craters in the Grieve data set are likely to have been the result of impacts by asteroids of non-cometary origin. However, at present, it is not possible to distinguish between the two types of craters. In any case, non-cometary impacts should be randomly distributed in time and hence should only produce a non-periodic background contribution, like that from

Table 3 Summary of the mathematical solutions for terrestrial and galactic periodicities

Episodes	Ref.	Most recent epoch (Myr BP)	Cycle (Myr)
Terrestrial			
Mass extinctions ($> 10\%$)	1	10 ± 7	30 ± 1
Impacts	49	$5 \pm 6^*$	$31 \pm 1^*$
Galactic			
Galactic plane crossings	13	~ 0	$33 \pm 3^*$

Terrestrial dates have been revised after ref. 5. Values for the most recent epoch are not significantly different from 0, in view of the size of the associated errors.

* Estimated error.

the occasional cometary impacts that do not originate from comet showers.

The distribution of the younger crater ages suggests that multiple impacts occurred around 15, 38, 65 and 100 Myr ago, as Seyfert and Sirkin⁵⁰ originally proposed. In view of the known dating uncertainties, the dates of these impact episodes are in reasonably good agreement with the dates of the mass extinction episodes of Table 1.

Comet or asteroid impacts have already been inferred from the discovery of geochemical anomalies occurring at the time of the late Cretaceous (66 Myr) extinctions^{51,52}, which included the disappearance of the dinosaurs, and at the time of the late Eocene (37 Myr) extinctions^{53,54}. Several microtektite layers have also been discovered in the 10-Myr interval between 40 and 31 Myr ago⁵⁵. Similar geochemical anomalies and microtektite layers should exist at other relevant boundaries.

If the geological consequences of episodic impacts go beyond immediate effects, as some authors have suggested⁴⁴⁻⁴⁷, we predict that other kinds of terrestrial data should also show a mean periodicity of ~ 30 Myr. With similar techniques, a periodicity of 34 Myr has recently been found in the pattern of geomagnetic reversals⁵⁶. Furthermore, we predict that when better extinction data for Palaeozoic times become available, the dominant cycle of major mass extinctions in these earlier times should be ~ 30 Myr. More speculatively, if one accepts our hypothesis that the basic pacemaker for biological extinctions is the Sun's vertical oscillation through the Galaxy, our line of reasoning can be reversed and we may refine the dynamical half-period of galactoververtical motion near the galactic plane by using the terrestrial results summarized in Table 3. We then find $P_{1/2} = 30 \pm 3$ Myr (estimated uncertainty). This seems to support a large local dynamical mass density of $\rho_0 \approx 0.2 M_{\odot} \text{ pc}^{-3}$. Whether this in turn implies the existence of 'missing matter' in the solar vicinity^{19,20} is still uncertain. A last point worth considering for its philosophical implications is that the present mechanism should be operating at roughly the same driving frequency throughout the galactic disk.

We thank D. M. Raup for supplying a preprint and for a valuable discussion, S. Lebedeff for translating an article, and E. M. Shoemaker for a critical review. Our work was supported in part by a NASA Cooperative Agreement (NCC 5-16, Task III) with Columbia University.

Received 15 November 1983; accepted 20 March 1984.

1. Raup, D. M. & Sepkoski, J. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 801–805 (1984).
2. Stothers, R. *Astr. Astrophys.* **77**, 121–127 (1979).
3. Fischer, A. G. & Arthur, M. A. *Soc. Econ. Paleont. Miner. Spec. Publ.* **25**, 19–50 (1977).
4. Raup, D. M. *Geol. Soc. Am. Spec. Pap.* **190**, 277–281 (1982).
5. Palmer, A. R. *Geology* **11**, 503–504 (1983).
6. McCrea, W. H. *Proc. R. Soc. A* **375**, 1–41 (1981).
7. Dilke, F. W. W. & Gough, D. O. *Nature* **240**, 262–294 (1972).
8. Hoyle, F. & Lyttleton, R. A. *Proc. Camb. phil. Soc.* **35**, 405–415 (1939).
9. Shapley, H. *Sky Telesc.* **9**, 36–37 (1949).
10. Steiner, J. & Grillmair, E. *Bull. geol. Soc. Am.* **84**, 1003–1018 (1973).
11. Williams, G. E. *Earth planet. Sci. Lett.* **26**, 361–369 (1975).
12. McCrea, W. H. *Nature* **255**, 607–609 (1975).
13. Innanen, K. A., Patrick, A. T. & Duley, W. W. *Astrophys. Space Sci.* **57**, 511–515 (1978).
14. Tamrazyan, G. P. *Izv. Acad. Nauk Azerb. SSR* **12**, 85–115 (1957).
15. Hatfield, C. B. & Camp, C. J. *Bull. geol. Soc. Am.* **81**, 911–914 (1970).
16. Meyerhoff, A. A. *Mem. Can. Soc. petrol. Geol.* **2**, 745–758 (1973).

17. Oort, J. H. in *Galactic Structure* (eds Blaauw, A. & Schmidt, M.) 455–511 (University of Chicago, 1965).
18. Eggen, O. J. *Publ. astr. Soc. Pacif.* **81**, 741–753 (1969).
19. Krisciunas, K. *Astr. J.* **82**, 195–197 (1977).
20. Hill, G., Hilditch, R. W. & Barnes, J. V. *Mon. Not. R. astr. Soc.* **186**, 813–830 (1979).
21. Rohlf, K. & Kreitschmann, J. *Astrophys. Space Sci.* **79**, 289–319 (1981).
22. Klugh, H. E. *Statistics, the Essentials for Research*, Ch. 10 (Wiley, New York, 1970).
23. Knude, J. *Astr. Astrophys.* **126**, 89–93 (1983).
24. Talbot, R. J. Jr. & Newman, M. J. *Astrophys. J. Suppl.* **34**, 295–308 (1977).
25. Ilovaisky, S. A. & Lequeux, J. *Astr. Astrophys.* **18**, 169–185 (1972).
26. Clark, D. H. & Caswell, J. L. *Mon. Not. R. astr. Soc.* **174**, 267–305 (1976).
27. Terry, K. D. & Tucker, W. H. *Science* **159**, 421–423 (1968).
28. Ruderman, M. A. *Science* **184**, 1079–1081 (1974).
29. Whitten, R. C., Cuzzi, J., Borucki, W. J. & Wolfe, J. H. *Nature* **263**, 398–400 (1976).
30. Clark, D. H., McCrea, W. H. & Stephenson, F. R. *Nature* **265**, 318–319 (1977).
31. Stark, A. A. in *Kinematics, Dynamics and Structure of the Milky Way* (ed. Shuter, W. L. H.) 127–133 (Reidel, Dordrecht, 1983).
32. Clube, S. V. M. & Napier, W. M. *Q. Jl. R. astr. Soc.* **23**, 45–66 (1982).
33. van den Bergh, S. J. *R. astr. Soc. Can.* **76**, 303–308 (1982).
34. Sanders, D. B., Scoville, N. Z. & Solomon, P. M. Preprint, Univ. Massachusetts (1984).
35. Cohen, R. S., Cong, H., Dame, T. M. & Thaddeus, P. *Astrophys. J.* **239**, L53–56 (1980).
36. Chandrasekhar, S. *Principles of Stellar Dynamics*, 190 (University of Chicago, 1942).
37. Vidal-Madjar, A., Laurent, C., Bruston, P. & Audouze, J. *Astrophys. J.* **223**, 589–600 (1978).
38. Begelman, M. C. & Rees, M. J. *Nature* **261**, 298–299 (1976).
39. Talbot, R. J. Jr., Butler, D. M. & Newman, M. J. *Nature* **262**, 561–563 (1976).
40. Butler, D. M., Newman, M. J. & Talbot, R. J. Jr. *Science* **201**, 522–525 (1978).
41. McKay, C. P. & Thomas, G. E. *Geophys. Res. Lett.* **5**, 215–218 (1978).
42. Hills, J. G. *Astr. J.* **86**, 1730–1740 (1981).
43. Oort, J. H. *Bull. astr. Inst. Neth.* **11**, 91–110 (1950).
44. Urey, H. C. *Nature* **242**, 32–33 (1973).
45. Napier, W. M. & Clube, S. V. M. *Nature* **282**, 455–459 (1979).
46. Clube, S. V. M. & Napier, W. M. *Earth planet. Sci. Lett.* **57**, 251–262 (1982).
47. Silver, L. T. & Schultz, P. H. (eds) *Geol. Soc. Am. Spec. Pap.* **190**, 1–528 (1982).
48. Pollack, J. B., Toon, O. B., Ackerman, T. P., McKay, C. P. & Turco, R. P. *Science* **219**, 287–289 (1983).
49. Grieve, R. A. F. *Geol. Soc. Am. Spec. Pap.* **190**, 25–37 (1982).
50. Seyfert, C. K. & Sirkis, L. A. *Earth History and Plate Tectonics*, 383–389 (Harper & Row, New York, 1979).
51. Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. *Science* **208**, 1095–1105 (1980).
52. Hsü, K. J. *Nature* **285**, 201–203 (1980).
53. Ganapathy, R. *Science* **216**, 885–886 (1982).
54. Alvarez, W., Asaro, F., Michel, H. V. & Alvarez, L. W. *Science* **216**, 886–888 (1982).
55. Keller, G., D'Hondt, S. & Vallier, T. L. *Science* **221**, 150–152 (1983).
56. Negi, J. G. & Tiwari, R. K. *Geophys. Res. Lett.* **10**, 713–716 (1983).

Periodic mass extinctions and the Sun's oscillation about the galactic plane

Richard D. Schwartz & Philip B. James

Department of Physics, University of Missouri, St Louis, Missouri 63121, USA

Raup and Sepkoski¹ have recently reported evidence for a 26-Myr periodicity in the occurrence of mass extinctions based on a study of marine fossils. The data baseline of 250 Myr suggests events of variable amplitude, with some of the strongest peaks associated with boundaries between major geological periods which have been defined by previous palaeontological studies. In a more limited quantitative study, Fischer and Arthur² earlier cited evidence for a 32-Myr period of major extinction events. Hatfield and Camp³ were among the first to suggest that mass extinctions might be correlated with periodic galactic phenomena, noting intervals of 80–90 Myr between major mass extinctions with an exceptionally strong mass extinction every 225–275 Myr. Here we point out a possible correlation between the 26-Myr extinction period and the Sun's oscillation about the galactic plane.

Due to the disk-like mass distribution of the Galaxy, stars in the disk (including the Sun) experience a restoring force towards the galactic plane which gives rise to harmonic oscillation perpendicular to the galactic plane. Stellar velocity distributions have been used empirically to derive the force law in the galactic plane⁴. Oort calculated a period of the Sun's 'z-oscillation' on the order of 68 Myr and an amplitude of ~100 pc on either side of the midplane of the Galaxy^{5,6}. More recent work by Bahcall⁷ indicates a slightly stronger force law in the galactic plane than that derived by Oort, suggesting an oscillation period of ~62 Myr. Hatfield and Camp³ cited earlier work of Trumpler and Weaver⁸ which suggested a period of 80 Myr for the oscillation, close to one of the major mass extinction intervals. In

addition, they attributed the strong mass extinctions at ~250-Myr intervals to the period of the Sun's revolution around the centre of the Galaxy.

In considering the Sun's harmonic oscillation about the galactic plane, it is important to recognize that the Sun crosses midplane twice in each period, and similarly that it reaches maximum distances from midplane twice each period. Thus any correlation relating to the Sun's position in this oscillation might well involve half the oscillation period, or ~31 Myr. The discrepancy of 5 Myr with the Raup and Sepkoski figure¹ may not be significant in view of the uncertainties associated with establishing each of the periods. In using the Harland time scale for their data set, Raup and Sepkoski recognized the difficulties in establishing an absolute time scale on the basis of a combination of chronostratigraphical sequences and radiometric dating. At the same time, the derivation of the force law in the galactic plane relies on assumptions involving isotropy of the stellar z-velocity distributions. Although isotropy may be indicated by z-velocity distributions of nearby catalogue stars within the local standard of rest (LSR), the LSR could itself possess a z-oscillation which, superposed on the solar oscillation, could yield an effective solar half-oscillation of <31 Myr. We suggest that present uncertainties would easily allow equal periods of ~30 Myr for the geological data as well as for the solar oscillation.

What circumstances could trigger the mass extinctions as a function of the Sun's galactic z-position? Passage of the Sun through dense dust clouds located at midplane could in principle provide a perturbing effect on the solar irradiation of Earth, thus initiating climatic changes which would affect the biosphere. However, there is no astronomical evidence for a widespread, flattened dust distribution of sufficient opacity to trigger such an effect with each solar midplane passage. Moreover, data suggest that the Sun is presently only ~8 pc north of the midplane of the Galaxy⁹. As the Sepkoski and Raup analysis indicates that we are presently about midway between mass extinction episodes (the two most recent episodes were ~11 Myr and 38 Myr in the past), this would suggest that the episodes are triggered as the Sun approaches its extreme positions away from the galactic plane. The exceptionally strong extinction event at 65 Myr (the Cretaceous-Tertiary boundary) could have resulted from the superposition of the impact of an asteroid on the longer time scale periodic extinction which was near maximum at the time. This would account for the fact that significant faunal extinctions in excess of those expected from the level of background extinction seem to have been well underway before the deposition of the iridium layer attributed to the impact^{10,11}.

What other agents of the galactic environment might serve to modulate evolution in the biosphere? Hatfield and Camp³ suggested that geomagnetic field reversals might be triggered by variations in the galactic magnetic field as the Solar System moves through the galactic plane. At times of null magnetic field strength during the reversals, the increased cosmic ray flux at the Earth's surface could imperil the biosphere. To our knowledge, however, there is no evidence for a periodicity of ~30 Myr in geomagnetic field reversals. Moreover, the galactic magnetic field is six orders of magnitude weaker than the geomagnetic field, and in view of the fact that the magnetic flux associated with the solar wind at Earth is greater than the galactic magnetic flux, it is difficult to understand how the galactic field could perturb the geomagnetic field.

There are two components of the galactic environment which might be expected to undergo substantial changes from the point of view of a star oscillating through a distance comparable to the scale height distribution of interstellar matter. First, electromagnetic radiation in the 2 keV–1 eV range is strongly absorbed by interstellar material which in general obscures sources located more than 1 kpc from the Sun. If a source of radiation is located at midplane at a distance D from the Sun, and if the extinction law can be written

$$\kappa_{\nu}(z) = \kappa_{\nu}(0) \exp(-z/L) \quad (1)$$

where L is the scale height of interstellar matter, the specific intensity of the source as observed from the Sun's present location near midplane ($z=0$) will be

$$I_\nu(0) = I_{0,\nu} \exp[-\kappa_\nu(0)D] \quad (2)$$

where $I_{0,\nu}$ is the specific intensity at the source. For a position one scale height above midplane ($z=L$), it can be shown that the observed intensity is approximately

$$I_\nu(L) \approx I_\nu(0) \exp[\kappa_\nu(0)D/e] \quad (3)$$

where we have used the condition $D \gg L$. Of special significance is the energy range 2–0.1 keV (soft X rays) where increasing flux is coupled with rapidly increasing optical depth. Using the cross-sections of Brown and Gould¹² and a column density of hydrogen of $6 \times 10^{22} \text{ cm}^{-2}$, characteristic, for example, of the distance to the galactic centre, optical depths are found to increase rapidly for photons with energy $< 1 \text{ keV}$: $\tau \approx 7$ at 1.5 keV, $\tau \approx 12$ at 1.0 keV, and $\tau \approx 42$ at 0.6 keV. Equation (3) indicates that an optical depth $[\kappa_\nu(0)D]$ of 20, for example, would lead to an enhancement by a factor of $\sim 1,600$ in the intensity observed at L in comparison with midplane. The integrated soft X-ray flux from the diffuse galactic X-ray background¹³, from discrete sources such as supernova remnants, and from the galactic centre¹⁴ could thus undergo a considerable increase as the Sun approaches a distance comparable to L ($\sim 125 \text{ pc}$)⁶ from midplane.

A second component of possible significance is the galactic cosmic-ray distribution. The distribution of cosmic rays with height above the galactic plane is unknown. However, if the cosmic rays are confined by the galactic magnetic field, and if the galactic magnetic field is tied to interstellar material, then cosmic rays should have a distribution similar to that of interstellar material. Thus at a scale height above the plane the cosmic-ray flux might be expected to be less than at midplane. Cosmic rays, like soft X rays, play a part in the ionization balance of the upper atmosphere, and in this case one might expect a decrease in the average ionization state as the Sun approaches the extrema of its oscillation. It is not clear whether changes in X-ray flux or changes in cosmic ray flux (which are both absorbed at high altitudes) would dominate.

Because of the harmonic nature of the oscillation, the Solar System would be expected to spend most of its time away from the midplane of the Galaxy. In particular, roughly 4.7 Myr would be spent within 10 pc of each extremum. Changes in cosmic rays in response to variations in solar activity are widely thought to correlate with terrestrial climate cycles^{15,16} although the mechanisms for such a relation are not understood. If such correlations are real, it is not unreasonable that long-term changes in these fluxes due to the Sun's z -oscillation could produce significant alterations in the biosphere.

We thank Drs Raup and Sepkoski for making a preprint of their work available, and for pointing out the earlier work of Hatfield and Camp.

Are periodic mass extinctions driven by a distant solar companion?

Daniel P. Whitmire

Department of Physics, University of Southwestern Louisiana, USL Box 44210, Lafayette, Louisiana 70504, USA

Albert A. Jackson IV

Computer Sciences Corporation, 1300 Bay Area Blvd, Houston, Texas 77058, USA

Raup and Sepkoski^{1–3} have recently reported evidence for a 26-Myr cycle in biological mass extinctions which, if real, requires an astronomical explanation. Here we investigate a model in which this extinction cycle is associated with the orbital period of a solar companion star. The required semi-major axis is (a_*), semi-major axis $\approx 8.8 \times 10^4 \text{ AU} = 1.4 \text{ light yr}$. Its highly eccentric orbit ($e_* \approx 0.9$) periodically brings the companion into the dense inner region ($\leq 2 \times 10^4 \text{ AU}$) of the comet cloud where it perturbs the orbits of large numbers of comets, initiating an intense comet shower⁴ in the Solar System which results in several terrestrial impacts over a period of 10^5 – 10^6 yr . The companion probably has a mass in the black dwarf range of 2×10^{-4} to $7 \times 10^{-2} M_\odot$, depending on its eccentricity and the density distribution of comets in the inner cloud, and is potentially observable in the infrared.

The putative extinction period, although uncertain by as much as $7 \times 10^6 \text{ yr}$ (ref. 1), places the companion roughly near aphelion at the present epoch. The aphelion distance is $Q_* = a_*(1 + e_*) \approx 2a_* \approx 1.8 \times 10^5 \text{ AU}$, which is near the outer boundary of the Oort comet cloud. Fossil evidence for the extinction cycle extends back $2.4 \times 10^8 \text{ yr}$ or 10 companion orbital periods. We consider first the long-term stability of the required orbit against perturbations due to the galactic nucleus, field stars and giant molecular clouds (GMCs).

The time-averaged separation radius is $\langle r_* \rangle = a_*(1 + \frac{1}{2}e_*^2) \approx 1.5a_* \approx 1.3 \times 10^5 \text{ AU}$, which is compatible with the limit of $\approx 2 \times 10^5 \text{ AU}$ imposed by the galactic field perturbation⁵. The time scale for the dissolution of the solar-companion binary due to stellar encounters is⁶

$$t \sim 1.7 \times 10^{15} \left(\frac{M_\odot + M_*}{a} \right) \text{ yr} \quad (1)$$

where the stellar masses are in solar units and the semi-major axis is in AU. Assuming the companion's mass M_* is $\ll M_\odot$ and setting $a = 8.8 \times 10^4 \text{ AU}$ yields $t = 2 \times 10^{10} \text{ yr}$. A somewhat shorter time scale is obtained using the mean specific energy transfer rate of $\sim 10^{-13} \text{ m}^2 \text{ s}^{-3}$ at 10^5 AU given by Bailey⁷.

Although the companion's energy and hence semi-major axis and orbital period are not expected to change significantly over an interval of $\sim 10^9 \text{ yr}$, the eccentricity and perihelion distance $q_* = a_*(1 - e_*)$ are more easily altered by stellar perturbations. In the context of a model for the density distribution of comets in the inner cloud (discussed below) the maximum perihelion distance is constrained by the required number of shower comets to be $\leq 8.8 \times 10^3 \text{ AU}$, corresponding to $e_* \geq 0.9$. Beginning with $e_* \approx 1$, the change in perihelion distance in successive orbits is a random walk process with secularly increasing mean perihelia. In an equivalent problem, Fernandez⁸ has calculated the mean eccentricity $\bar{e}$ of comet orbits as a function of their semi-major axis (assumed approximately unchanged) for several different time intervals. Beginning with an initial eccentricity $e_0 \approx 1$, Fernandez finds that $\bar{e}$ is essentially unchanged over 10^8 yr independent of the semi-major axis, but significantly changed ($\bar{e} \approx 0.75$) over 10^9 yr for $a \approx 8.8 \times 10^4 \text{ AU}$. These results depend on the r.m.s. change in angular momentum (or velocity) per orbit and various estimates^{7,9,10} differ by as much as an order of magnitude. In view of this uncertainty, we conclude that the companion's eccentricity need not necessarily change appreciably (that is,

Received 16 November 1983; accepted 8 March 1984.

1. Raup, D. M. & Sepkoski, J. J. Jr *Proc. natn. Acad. Sci. U.S.A.* **81**, 801–805 (1984).
2. Fischer, A. G. & Arthur, M. A. *Soc. Econ. Paleont. Miner. Spec. Publ.* **25**, 19–50 (1977).
3. Hatfield, C. B. & Camp, M. J. *Bull. geol. Soc. Am.* **81**, 911–914 (1970).
4. Oort, J. H. *Bull. astr. Inst. Neth.* **15**, 45 (1960).
5. Oort, J. H. in *Stars and Stellar Systems* Vol. 5 (University of Chicago Press, Chicago, 1965).
6. Mihalas, D. & Routly, P. M. *Galactic Astronomy*, 221 (Freeman, San Francisco, 1968).
7. Bahcall, J. N. *Astrophys. J.* **276**, 169 (1984).
8. Trumpler, R. J. & Weaver, H. F. *Statistical Astronomy* (Dover, New York, 1962).
9. Allen, C. W. *Astrophysical Quantities*, 283 (Athlone, London, 1973).
10. Pollack, J. B., Toon, O. B., Ackerman, T. P. & McKay, C. P. *Science* **219**, 287–289 (1983).
11. Kauffmann, E.-G. in *Proc. Conf. on Large Body Impacts and Terrestrial Evolution*, 23 (Lunar and Planetary Institute, Snowbird, Utah, 1981).
12. Brown, R. L. & Gould, R. J. *Phys. Rev. D* **1**, 2252–2256 (1970).
13. Nousek, J. A., Fried, P. M., Sanders, W. T. & Krausharr, W. L. *Astrophys. J.* **258**, 83–95 (1982).
14. Watson, M. G., Willingale, R., Grindlay, J. E. & Hertz, P. *Astrophys. J.* **250**, 142–154 (1981).
15. Damon, P. E., Lerman, J. C. & Long, A. A. *Rev. Earth planet. Sci.* **6**, 457–494 (1978).
16. Eddy, J. A. *Science* **192**, 1189–1202 (1976).

$\Delta e \leq 0.1$) due to stellar perturbations over 2.4×10^8 yr although a significant decrease would be expected over the age of the Solar System.

The most serious long-term threat to the companion's stability may be perturbations due to the passage of giant molecular clouds of mass $\sim 10^5$ – $10^6 M_\odot$ within a few tens of parsecs of the Sun every $\sim 10^8$ yr (ref. 7), which is comparable to the required lifetime of the present orbit. Comets (and a low-mass companion) at $r \geq 2 \times 10^4$ AU may be disassociated from the Sun by such an encounter. Thus, before the last encounter with a GMC, the present Oort cloud comets (and companion) must have been in closer solar orbits^{4,7}. (This assumes that comets and the companion are primordial and have not been captured from interstellar space or from GMCs^{11,12}.) To have survived this encounter, the companion's semi-major axis must previously have been $< 8.8 \times 10^4$ AU. In connection with this evolutionary scenario, Bahcall and Soneira¹³ found that $\approx 15\%$ of the stars they surveyed were binaries having projected separations of the order of 2×10^4 AU. The apparent absence of significant numbers of binaries with projected separations greater than this safe limit may be explained by encounters with GMCs over the lifetime of the binaries⁴. Our hypothesized solar companion may have been in a similar orbit at $\approx 8.8 \times 10^4$ AU during most of the lifetime of the Solar System. The last close encounter with a GMC did not completely disassociate the companion from the Sun but left it in its present eccentric and loosely bound orbit. Alternatively, though less likely, the companion may have been captured during, or subsequent to, the last encounter with a GMC.

The occurrence of a brief but intense comet shower due to the rare (and random) passage of a field star within 2×10^4 AU of the Sun has previously been investigated by Hills⁴. The existence of a dense inner core of comets within this distance has been proposed as a source to replenish comets ejected by encounters with GMCs^{4,7,14}. Comets from the dense inner cloud are normally unobserved because distant stellar perturbations are not effective in randomizing their orbits and maintaining a filled loss cone. Hills estimates that the number of comets in the inner cloud may be several orders of magnitude more than the number in the observed Oort halo ($\sim 10^{11}$). A solar mass field star passing the Sun with impact parameter b can randomize comets with semi-major axes $a \geq b$. Once randomized, the fraction of comets having semi-major axes a and perihelia q is $\approx 2q/a$, if $q \ll a$.

In the numerical example considered by Hills, a $1 M_\odot$ star of velocity 30 km s^{-1} passes the Sun at $b = 3 \times 10^3$ AU. (The mean interval between such encounters is 5×10^8 yr.) The resulting average comet flux in the inner Solar System ($q \leq 2$ AU) was estimated to be $10^4 \text{ yr}^{-1} \approx 1 \text{ h}^{-1}$. The duration of this shower was 7×10^5 yr and the total number of shower comets was $\sim 7 \times 10^9$. The corresponding expected number of terrestrial impacts is $\sim 4 \times 7 \times 10^9 P_i$ where P_i is the impact probability per perihelion passage for a random long-period Earth-crossing comet which makes on average four perihelion passages⁵. According to Weissman¹⁵, $P_i \sim 2 \times 10^{-9}$ and we could therefore expect ~ 50 terrestrial impacts from this shower. Of course, the poorly known mass distribution of comets has not been taken into account and the estimated impact number would be accordingly scaled down for massive comets (producing craters ≥ 100 km, say) and scaled up for less massive comets (producing craters ≤ 10 km, say). As noted by Hills, the resulting environmental stress may be expected to significantly affect the distribution of species.

Hills' impulse analysis cannot be taken over directly in the case of a low-mass bound companion. Here we consider the limit of the two-body problem of the scattering of comets by the companion and neglect the relatively small acceleration given the Sun by the companion. The two-body approximation is valid for the scattering of comets which come within the companion's sphere of action radius $d = rM_*^{1/3}$, where r is the Sun-companion separation and M_* is the mass of the companion in solar units. The fraction of comets within this distance

which will be scattered into orbits with perihelia $\leq q$ is $\approx (2q/r)$. The rate at which these comets are scattered is

$$\dot{N} \approx n\sigma u \left(\frac{2q}{r} \right) \quad (2)$$

Here $n = n(r)$ is the number density of comets, $\sigma = \pi d^2$ is the effective total two-body scattering cross-section and u is the relative velocity between the companion and comet.

The number density of comets can be expressed as $n(r) = C(p)r^{p-4}$ where p is the model index¹⁴. In the following analysis we adopt the index $p = -1/2$ which is equivalent to the power law model used by Hills⁴ and discussed by Bailey¹⁴. In this model $n(r) \approx 2 \times 10^{17} r^{-4.5} \text{ AU}^{-3}$, where r is in AU. The relative velocity u is related to the heliocentric circular velocity at r by $u = Uv$, where v is in units of AU yr^{-1} and U is a geometrical factor which depends on (a_*/r) , the eccentricities of the companion and comet orbits, and on the relative inclinations of their orbital planes¹⁶. For all encounter radii and eccentricities of interest, $U \approx 2$. Making these substitutions in equation (2) gives

$$\dot{N} \approx 6.3 \times 10^{19} q M_*^{2/3} r^{-4} \quad \text{comets yr}^{-1} \quad (3)$$

where q and r are in AU.

Integrating equation (3) around an elliptical orbit yields

$$N \approx 3.2 \times 10^{19} q M_*^{2/3} a_*^{-5/2} f(e_*) \quad \text{comets per orbit} \quad (4)$$

where q and a_* are in AU and

$$f(e_*) = \frac{2 + e_*^2}{(1 - e_*^2)^{5/2}} \quad (5)$$

Setting $q = 1$ AU, $a_* = 8.8 \times 10^4$ AU and $N \geq 5 \times 10^8$ gives the required companion mass as a function of eccentricity

$$M_* \geq 2.2 \times 10^2 f(e_*)^{-3/2} \quad (6)$$

The fact that the companion has not yet been identified suggests that its mass may be below the minimum mass of $\approx 0.07 M_\odot$ required for hydrogen ignition. Inserting this mass into equation (6) gives $e_* \geq 0.9$. A rough lower mass limit can be estimated by imposing the condition that the total mass of comets in the inner cloud be $\leq 300 M_\oplus$ or $\leq 2 \times 10^{13}$ comets, assuming an average comet mass of 10^{17} g. This limit has recently been inferred by Bailey on the basis of IR observations¹⁷. In the context of our model for $n(r)$, the total number of comets in the inner cloud is $\approx 1.7 \times 10^{18} r_{\min}^{-3/2}$, where $r_{\min}$ (in AU) is the effective inner boundary of our model cloud. Setting the total number of comets $= 2 \times 10^{13}$ gives $r_{\min} \approx 2 \times 10^3$ AU, which is the value suggested by Hills⁴. We now require $q_* \geq r_{\min}$ or $e_* \leq 0.98$. Inserting this eccentricity into equation (6) yields a lower limit of $M_* \geq 2.4 \times 10^{-4}$, which is somewhat less than the mass of Saturn.

The fossil record requires that the duration of the shower be less than a few million years. The strong radial dependency of $\dot{N}$ in our model means that there will be a non-negligible contribution to the shower only near perihelion passage and for all perihelia of interest the characteristic time scale is $\leq 10^6$ yr (including four return passages). Other less condensed comet cloud models (that is, $p > 0$) were found to be inconsistent with either the required number of comets or the shower duration time scale for $M_* \leq 0.07$.

As previously noted, if the time scale for encounters with GMCs is less than the age of the Solar System the companion must have been in a closer orbit at an earlier epoch. If the earlier orbit was approximately circular at 2×10^4 AU, say, then for $M_* \leq 0.07$ and $q = q_{\text{loss}} = 15$ AU, the number of comets lost per orbit would have been $\leq 10^9$. Assuming that the companion was in this orbit for $\sim 4 \times 10^9$ yr, the total number of comets lost to 15 AU would have been $\leq 10\%$ of the estimated current number of $\sim 2 \times 10^{13}$ assuming $p = -1/2$. We note that a companion-induced asymmetry in the inner comet cloud might be of interest in connection with the well known anomalous residuals in the motion of Neptune, as discussed recently by Bailey in the context

of a symmetrical comet shell model¹⁴. The contribution to terrestrial cratering while in this orbit is found from equation (3) with $q = 1$ AU. For $M_* \leq 0.07$, $\dot{N} \approx 66$ comets yr^{-1} , which is not significantly greater than various estimates of the current flux. These crude estimates are model-dependent and are only intended to show that the suggested evolutionary scenario is not necessarily in conflict with the present existence of the inner cloud and the cratering record. A detailed analysis of the evolution of the comet cloud plus companion is beyond the intent of this paper.

As noted above in connection with Hill's analysis, the intrinsic mass distribution of comets is uncertain. Density functions implicitly refer to observable comets of average size since normalization is based on the estimated number of observed long-period comets per year. Here we assume (as did Hills) that the relevant average-size comet has a radius of ~ 1 km and that the estimate of $\sim 5 \times 10^8$ comets per impact applies to this average comet. Qualitatively, we expect that there are intrinsically many more comets of smaller size. For every 1 km comet that strikes the Earth there will statistically be numerous smaller impacts, as observed in the cratering record. A large 10 km comet will similarly require on average more than $\sim 5 \times 10^8$ comets per impact. We thus anticipate a variation from cycle to cycle in the magnitude of the biological extinctions, depending on the maximum size impact comet as well as variations in the companion's perihelia. However, each cycle could always produce a number of collisions with ≤ 1 km comets. The size and number of comets necessary to significantly perturb the biosphere is uncertain¹⁵, although in the case of the major Cretaceous-Tertiary extinction, estimates suggest that the impacting body or bodies had the equivalent mass of a single ≈ 10 km asteroid or comet¹⁸.

A black dwarf at a distance of $\langle r_* \rangle \approx 1.3 \times 10^5$ AU is potentially observable in the IR (for example, see refs 19–21). Such objects evolve along a vertical Hyashi track in the HR diagram until a final radius of $\sim 0.1 R_\odot$ is reached, after which they cool radiatively²². For example, a dwarf of mass $0.02 M_\odot$ and age 5×10^9 yr has an absolute $2.2\text{--}\mu\text{m}$ magnitude of 21.3 (refs 19, 22) and a corresponding apparent magnitude of 15.3 at 1.3×10^5 AU. The entire black dwarf mass range could similarly be observed at this distance. A recent search for black dwarfs around nearby stars concluded that these objects must be rare as companions¹⁹. However, the area searched around the candidate stars was 65×65 arc s or typically $\leq 10^3$ AU and more distant dwarfs (as well as dwarfs at distances ≤ 10 arc s) were not excluded by this search. Although the companion's orbital period, semi-major axis, eccentricity, present distance and mass are all reasonably bracketed by the model, its orbital inclination and aphelion direction are unknown. Nonetheless, the companion is expected to be sufficiently bright in the IR for it to be observable in a full sky survey at the appropriate wavelength. The recent IRAS survey has thus far not found a tenth planet or solar companion, however most of the data analysis is still in progress²³.

The 2.6×10^7 yr extinction period has not as yet been detected in the cratering record (see below) on the Earth or Moon, although there is some evidence for a mean interval of 5×10^7 yr between major impacts^{2,3,24}. Finally, we note that the various arguments^{20,25–29} invoked to exclude the existence of an hypothesized³⁰ solar companion at $\sim 10^3$ AU are not applicable to the much more distant low-mass companion considered here.

Since submission of this paper we have received a preprint from Alvarez and Muller³¹ reporting the discovery of a periodicity in the terrestrial cratering rate with a period and phase that closely match the extinction period. The cratering period has subsequently been confirmed in an independent analysis by D. Raup and J. Sepkoski (personal communication).

We thank Dr John J. Matese for helpful discussions and Dr M. E. Bailey who made several valuable suggestions.

4. Hills, J. G. *Astr. J.* **86**, 1730–1740 (1981).
5. Wyatt, S. P. & Faintich, M. B. *Bull. Am. astr. Soc.* **3**, 368 (1971).
6. Heggie, D. C. *Mon. Not. R. astr. Soc.* **173**, 729–787 (1975).
7. Bailey, M. E. *Mon. Not. R. astr. Soc.* **204**, 603–633 (1983).
8. Fernandez, J. A. *Icarus* **42**, 406–421 (1980).
9. Weissman, P. R. *Nature* **288**, 242–243 (1980).
10. Öpik, E. J., *Astrophys. Space Sci.* **21**, 307–398 (1973).
11. Clube, S. V. M. & Napier, W. M. *Q. Jl R. astr. Soc.* **23**, 45–66 (1982).
12. Napier, W. N. & Staniucha, M. *Mon. Not. R. astr. Soc.* **198**, 723–735 (1982).
13. Bahcall, J. N. & Soneira, R. M. *Astrophys. J.* **246**, 122–135 (1981).
14. Bailey, M. E. *Nature* **302**, 399–400 (1983).
15. Weissman, P. R. *Geol. Soc. Am. Spec. Pap.* **190**, 15–24 (1982).
16. Öpik, E. *Interplanetary Encounters*, Ch. 1 (Elsevier, Amsterdam, 1976).
17. Bailey, M. E. *Mon. Not. R. astr. Soc.* **205**, 47p–52p (1983); *Mon. Not. R. astr. Soc.* (in the press).
18. Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. *Science* **208**, 1095–1108 (1980).
19. Jameson, R. F., Sherrington, M. R. & Giles, A. B. *Mon. Not. R. astr. Soc.* **205**, 39p–41p (1983).
20. Heinrichs, H. F. & Staller, R. F. A. *Nature* **273**, 132–134 (1978).
21. Reynolds, R. T., Tarter, J. C. & Walker, R. G. *Icarus* **44**, 722–779 (1980).
22. Stevenson, D. J. *Proc. astr. Soc. Aust.* **3**, 227 (1978).
23. Eberhart, J. *Sci. News* **124**, 324 (1983).
24. Shoemaker, E. in *Meet. on Dynamics of Extinctions* (Northern Arizona University, 1983).
25. Kirk, J. *Nature* **274**, 667–669 (1978).
26. Pineault, S. *Nature* **275**, 727–730 (1978).
27. Wilkins, D. *Nature* **282**, 696–697 (1979).
28. Kirk, J. *Nature* **286**, 306 (1980).
29. Wilkins, D. *Nature* **286**, 306 (1980); *Astr. Astrophys.* **98**, 30–33 (1981).
30. Harrison, E. R. *Nature* **270**, 324–326 (1977).
31. Alvarez, W. & Muller, R. *Nature* **308**, 718–720 (1984).

Extinction of species by periodic comet showers

Marc Davis*, Piet Hut† & Richard A. Muller‡

* Departments of Astronomy and Physics, University of California, Berkeley, California 94720, USA

† Institute for Advanced Study, Princeton, New Jersey 08540, USA

‡ Department of Physics and Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, USA

A 26-Myr periodicity has recently been seen in the fossil record of extinction in the geological past¹. At least two of these extinctions are known to be associated with the impact on the Earth of a comet or asteroid with a diameter of a few kilometres (refs 2, 3). We propose that the periodic events are triggered by an unseen companion to the Sun, travelling in a moderately eccentric orbit, which at its closest approach (perihelion) passes through the 'Oort cloud' of comets which surrounds the Sun (ref. 4; see ref. 5 for a review and ref. 6 for a more recent analysis). During each passage this unseen solar companion perturbs the orbits of these comets, sending a large number of them (over 1×10^9) into paths which reach the inner Solar System. Several of these hit the Earth, on average, in the following million years. At present the unseen companion should be approximately at its maximum distance from the Sun, ~ 2.4 light yr, and it will present no danger to the Earth until approximately AD 15,000,000.

The possibility that major extinctions occurred in a periodic manner was suggested by Fischer and Arthur⁷, who believed that the periodicity was driven by a terrestrial mechanism, but it was only the detailed statistical analysis of Raup and Sepkoski¹ that eliminated many questions regarding systematic bias. They weighted each of the 39 customary stratigraphical stages according to the percentage of families in the preceding stage that were extinct in the following stage. A 26-Myr period is evident in the weighted data, and this period was shown by Fourier analysis and Monte Carlo simulation to be statistically significant with a confidence level better than 99%. The boundaries separating stages with large family extinctions are termed 'extinction events', although there is no direct evidence that the extinctions actually occurred at these boundaries. Large excesses of iridium had been found near two of these boundaries by Alvarez *et al.*^{2,3}, who concluded that the events were triggered by the impact of an asteroid or comet of a few kilometres diameter. Such impacts throw a large amount of dust into the atmosphere, and the extinctions could have been caused by the resulting lack of sunlight, temporary changes in climate, acidification of rain, or by some combination of such factors^{8,9}.

Received 3 January; accepted 16 March 1984.

1. Raup, D. & Sepkoski, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 801–805 (1984).

2. Simon, C. *Science News* **124**, 212 (1983).

3. Lewin, R. *Science* **221**, 935–937 (1983).

While impacts on the Earth by asteroids or comets provide plausible explanations for the individual events in which many species are extinguished, it is difficult to explain the precise periodicity of such collisions. Napier and Clube¹⁰ had proposed that periodic catastrophes could be triggered by the capture of comets as the Sun passed through the spiral arms of the Galaxy (for a more recent discussion, see ref. 11). However, it is difficult to reconcile the Napier–Clube model with the relatively stable periodicity discovered by Raup and Sepkoski, in which the last four of the extinction events occurred within 2 Myr of the time predicted by an exact 26-Myr cycle. In addition, as pointed out to us by the Alvarez group, their measurements of isotope ratios of iridium and rhenium imply that the material that hit the Earth was of Solar System origin. The oscillations of the Sun in and out of the plane of the Galaxy have a half-period that roughly matches that required, and one might hypothesize, for example, the presence of an extremely thin layer of debris in the galactic plane that is encountered by the Solar System on each passage. This has the same difficulty as the Napier–Clube model; it predicts impacts with extra-Solar System material. In addition, the Sun is presently near the galactic plane, moving upwards with the relatively high velocity of $\sim 6 \text{ km s}^{-1}$. As the last extinction event occurred 11 Myr ago, we are almost halfway between extinctions; thus we have the wrong phase for such an explanation.

If we try to account for the periodicity by postulating an object that orbits the Sun and comes close to the Earth every 26 Myr, perturbations from nearby stars will prevent the orbit from coming into the inner Solar System more than once. An object with this period has a large semi-major axis of $\sim 88,000 \text{ AU}$ (where 1 AU is the mean Earth–Sun separation of $1.5 \times 10^{13} \text{ cm} = 1.6 \times 10^{-5}$ light years). If this object passes within 10 AU of the Sun, then its orbit is highly elliptical, with an eccentricity greater than 0.9999. An equivalent way of saying this is to note that the object has very low angular momentum—this not only requires an unlikely fine-tuning of the orbit, but it is also unstable. Within one orbital period the gravitational perturbations of passing stars will cause the orbit to gain enough angular momentum to increase the perihelion distance to more than 100 AU, virtually eliminating its direct effect on the inner Solar System.

However, an unseen solar companion need not come close to the Sun to perturb the cloud of comets that surrounds the Sun at large distances. Oort and others^{4,6} showed that there must be about 10^{11} comets in such a cloud with semi-major axis greater than or equal to 30,000 AU. There may be considerably more comets around the Sun than this, as comets with smaller orbits are not significantly perturbed by most passing stars and therefore would usually not be observed on Earth. Hills¹² has estimated the total number of comets to be closer to 10^{13} ; even so, the total mass of these comets is probably less than that of Jupiter. The unseen solar companion would perturb such smaller orbits in a periodic manner even if its perihelion were as large as 30,000 AU. With a semi-major axis of 10^5 AU , the orbital eccentricity, e , is 0.7 or greater. Such an orbit requires no ‘fine tuning’; in fact, one expects a distribution of eccentricities which uniformly fill phase space to be flat in e^2 , with the fraction of orbits having eccentricity greater than e given by¹³

$$N/N_0 = 1 - e^2 \quad (1)$$

Thus the r.m.s. value of a random distribution of binary orbits has a mean value for the eccentricity of $(0.5)^{1/2} = 0.7$, and this eccentricity is adequate for our orbit. Of course, the larger the eccentricity, the shorter the duration of close passage to the Sun and the shorter will be the periods of maximum perturbation. However, no very tight bounds have been derived from the fossil records on the precise localization in time of the extinction events, so eccentricities as small as 0.7 are still possible. Passing stars will still perturb the orbit of this companion, but they will only gradually change the period and the perihelion. The present orbit, with a semi-major axis of 10^5 AU , will be significantly disturbed and possibly disrupted on a time scale of

$\sim 2 \times 10^9 \text{ yr}$. (Different authors find slightly different values, between 1 and $4 \times 10^9 \text{ yr}$; refs 13–15. This orbit has probably evolved through a stochastic process, perturbed by random passing stars, from an initial orbit with a shorter period and smaller semi-major axis. Over the past $2 \times 10^8 \text{ yr}$ the semi-major axis may have increased by $\sim 10\%$, and the period by 15% (from Kepler’s third law) as the orbital energy decreases linearly on average^{13,14}. The uncertainties in the extinction dates are large enough to be compatible with such a drift in period. Once better data are available, it will be useful to test a parabolic fitting to the dates of major extinctions, although purely stochastic ‘jitter’ from individual stellar passages will always remain in the data even with absolutely perfect dating methods.

What effect will a single passage of a solar companion have on the comets in the Oort cloud? We will follow the detailed analysis of Hills¹² who considered the effects of random (non-periodic) stars passing close to the Sun. Normally the orbits of the comets are distributed isotropically within the cloud, except for orbits that enter the inner Solar System. Orbits which pass close enough to the Sun to be perturbed by Jupiter or Saturn (which have masses of order $10^{-3} M_\odot$) are swept out of the Oort cloud, either ejected into hyperbolic orbits or captured into short period (recurrent) orbits. The region in velocity space that is empty because of this effect is known as the ‘loss cone’, and it contains all the orbits which reach the inner Solar System from the Oort cloud. When a star or other massive object passes through the Oort cloud, the orbits of the comets will be perturbed and the loss cone will begin to fill. Hills showed that the fraction of the loss cone that will be filled is given by:

$$F = \left(\frac{27}{8}\right) \left(\frac{M^2}{M_\odot^2}\right) \left(\frac{a^4}{P^4}\right) \left(\frac{GM_\odot}{qv^2}\right) \quad (2)$$

where M is the mass of the perturbing star, v is its velocity (assumed by Hills to be $\sim 30 \text{ km s}^{-1}$), and P is its distance of closest approach to the Sun; $M_\odot$ is the mass of the Sun, a is the semi-major axis of the comets affected, q is the minimum distance from the Sun that the comet must reach (1 AU if it is to hit the Earth), and G is the gravitational constant. Normally the Earth sits in the quiet ‘eye’ of the storm of comets, and the comets we see are those that have been perturbed into this normally quiet loss cone region by randomly passing stars.

Hills analysed the particular case of a star with mass similar to that of the Sun passing within 3,000 AU of the Sun, an event that should occur roughly every 500 Myr. For this situation, each of the bracketed terms in equation (2) is of order unity, and the loss cone will be filled. In less than a hundred thousand years a ‘shower’ of 10^9 comets will reach the inner Solar System. Using the estimates of Weissman¹⁶ and Everhart¹⁷ for the probability of comets hitting the Earth, Hills concluded that for the duration of the shower (10^5 – 10^6 yr), between 10 and 200 comets will hit the Earth. He mentioned the possibility that such a shower, triggered by the rare passing star that comes within 3,000 AU of the Sun, could be responsible for the Cretaceous–Tertiary extinctions.

One can arrive at a value similar to that of Hills for the number of impacts on the Earth from a comet shower from the following simple considerations. For a comet moving in an ellipse of eccentricity e and semi-major axis a , the distance of closest approach q is given by

$$q = a(1 - e) \quad (3)$$

The fraction of comets with e between 1 and e is given by equation (2). Combining these two equations gives

$$N/N_0 = 1 - e^2 = 1 + (a - q)^2/a^2 \approx 2q/a$$

where we have neglected the term $(q/a)^2$. Most of the $N_0 = 10^{13}$ comets in the cloud will be between 10^3 and 10^4 AU . Taking $q = 1 \text{ AU}$ and $a = 10^4 \text{ AU}$, we find $N = 2 \times 10^9$ comets showering within the Earth’s orbit. (The number of comets that will reach Jupiter’s orbit at 5 AU is 10^{10} .) The probability that an individual comet will hit the Earth on a single pass is roughly the projected area of the Earth divided by the area of

its orbit, or 1.6×10^{-9} . Each comet will make, on average, four trips to the inner Solar System¹², and on each trip it has two opportunities to hit the Earth. Taking these values together, we find that the total number of comets expected to hit the Earth is $2 \times 10^9 \times 1.6 \times 10^{-9} \times 4 \times 2 = 25$. Of course this number depends directly on the number of comets in the inner part of the Oort cloud, for which we have no direct evidence.

Using equation (2) we now extrapolate from the case of a random passing star to the situation of a companion to the Sun passing within 30,000 AU at perihelion. The distance of closest approach P is now 10 times larger than for the random passing star case, but the velocity of a companion at this distance is $\sim 0.2 \text{ km s}^{-1}$, 150 times less than the velocity assumed by Hills. Thus, the loss cone will be filled for the same value of a , the comet semi-major axis, of 3,000 AU, and roughly the same number of comets will enter the inner Solar System and hit the Earth, that is, 10–200. Although the impulse approximation assumed by Hills breaks down here, the effect is still of the same order of magnitude. If the mass of the companion is much smaller than $0.1 M_{\odot}$, the loss cone will not be completely filled and only a few comets will hit the Earth. Of course, our estimates are approximate, and we probably know the number of comets in the Oort cloud only within one or two orders of magnitude. The calculations do show, however, that the model is plausible. Fluctuations in the eccentricity of the orbit (and hence in the perihelion distance) could conceivably account for the slow overall modulation seen in the intensity of the mass extinctions.

The major difficulty with our model is the apparent absence of an obvious companion to the Sun, and the existence of such an object is its most important prediction. We take this prediction seriously largely because of our inability to find any simpler explanation for the periodicity consistent with known facts. Unfortunately, the data are insufficient to allow us to determine where in the sky to look for the as yet unseen solar companion. The known stars nearest to the Sun have been discovered either because of their high apparent brightness, their large proper motion, or their association with other nearby stars. Unfortunately, our proposed companion star is likely to have none of these characteristics. In his review of the known nearby stars¹⁸, van de Kamp explicitly states that the existence of a "distant dwarf star companion (for the Sun) is not excluded".

Harrison¹⁹ considered the possibility of a nearby solar companion affecting apparent pulsar frequencies but his analysis indicated a star too close to the Sun to have the long period we require. (An additional analysis of Harrison's suggestion can be found in ref. 20.) If the solar companion is a black hole or a brown dwarf (a small star which never heated to ignition temperature) then it may be difficult to find. An intense X-ray and γ -ray source, Geminga, has been proposed as an object that may be very close to the Sun, although the present limits place it at less than 1,000 light yr²¹. If the companion is a hydrogen-burning M dwarf, its apparent magnitude will be between 4 and 12. There are over 10^6 stars in the sky within this range, and one of them may be the Sun's companion. The companion will have negligible radial velocity; from our estimate of its present distance (2.4 light yr, calculated assuming that the mass of the star is substantially less than that of the Sun) we expect an annual parallax motion of ± 1.4 arc s and a proper motion of 0.01 arc s per year. The large parallax is probably the key to finding the star. The parallax and proper motion are not large enough for the companion to have been spotted in full-sky surveys that use large proper motion to identify nearby stars²². It is possible that the companion may not have been identified as such even if it were as bright as 9th magnitude, and a careful search of star catalogues may provide some candidates, but we suspect that it

would have been noted years ago unless it is at the faint limit. Analysis of the IRAS data base may yield a candidate brown dwarf. Weakly bound binaries with a semi-major axis of 1.4 light yr are rare in the Galaxy. The observations show a steep drop-off above 0.7 light yr²³ and theoretical arguments (refs 13–15 and I. R. King and J. M. Retterer, in preparation) predict that separations greater than this should occur in fewer than 10^{-3} of star systems. It is possible that the conditions which drive evolution on the Earth are rarer in the Galaxy than had been supposed previously.

The number of comets that arrive in a single shower may be as much as one or two orders of magnitude greater than the number that arrive between showers. This has important implications for our understanding of Solar System physics and evolution. Since the comets fall in from many regions of the Oort cloud, their arrival in the vicinity of the Earth will be spread out over a considerable period of time, perhaps a million years or more. Thus, we do not expect the periodicity in comet impacts to be exact, but instead it should have a slight 'jitter' or variability of about a million years. One should be able to find evidence in the geological record, perhaps by looking for closely spaced iridium layers or by further studies of isotope ratios, that in the average extinction the Earth was hit by more than one comet.

If and when the companion is found, we suggest it be named Nemesis, after the Greek goddess who relentlessly persecutes the excessively rich, proud and powerful. We worry that if the companion is not found, this paper will be our nemesis.

After this paper was submitted, W. Alvarez and R.A.M.²⁴ found a periodicity in the ages of large impact craters on the Earth, with a period and phase that closely match those of the mass extinctions. In addition, we became aware of evidence in micro-tektite records²⁵ that there were at least two separate impacts near the Eocene–Oligocene boundary.

We thank numerous colleagues for critical discussion and helpful suggestions and, in particular, L. W. Alvarez, J. N. Bahcall, F. Crawford, J. Kare, I. King, H. Spinrad and S. Perlmuter. During this research R.A.M. was supported by a fellowship from the John D. and Catherine T. MacArthur Foundation. It was partially funded by the NSF Alan T. Waterman Award, by NSF grants EAR-81-15858 and PHY82-17352, and by the Department of Energy under contract DE-AC03-76SF00098.

Received 3 January; accepted 8 March 1984.

1. Raup, D. M. & Sepkoski, J. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 801–805 (1984).
2. Alvarez, L. W., Alvarez, W., Asaro, F., & Michel, H. V. *Science* **208**, 1095–1108 (1980).
3. Alvarez, W., Asaro, F., Michel, H. V. & Alvarez, L. W. *Science* **216**, 886–888 (1982).
4. Oort, J. H. *Bull. astr. Insts. Neth.* **11**, 91 (1950).
5. Oort, J. H. in *The Moon, Meteorites, and Comets*, 665 (eds Middlehurst, B. M. & Kuiper, G. P.) (University of Chicago Press, 1963).
6. Marsden, B. G. & Roemer, B. in *Comets* (ed. Wilkening, L. L.) 707–733 (University of Arizona Press, Tucson, 1982).
7. Fischer, A. G. & Arthur, M. A. *Soc. Econ. Paleont. Min. Spec. Publ.* **25**, 19–50 (1977).
8. Alvarez, L. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 627–642 (1983).
9. Silver, L. T. & Schultz, P. H. (eds) *Geol. Soc. Am. Spec. Pap.* 190 (Boulder, Colorado, 1982).
10. Napier, W. M. & Clube, S. V. M. *Nature* **282**, 455–459 (1979); *Earth planet Sci. Lett.* **57**, 251–262 (1982).
11. Shoemaker, E. M. *A. Rev. Earth Planet. Sci.* **11**, 461–494 (1983).
12. Hills, J. G. *Astron. J.* **86**, 1730–1740 (1981).
13. Heggie, D. C. *Rev. mex. Astr. Astrofis.* **3**, 169 (1977); *Mon. Not. R. astr. Soc.* **173**, 729–787 (1975).
14. Retterer, J. M. & King, I. R. *Astrophys. J.* **254**, 214–220 (1982).
15. Bailey, M. E. *Mon. Not. R. astr. Soc.* **204**, 603–633 (1983).
16. Weissman, P. R. *Astr. Astrophys.* **85**, 191–196 (1980).
17. Everhart, E. *Astr. J.* **74**, 735–750 (1969).
18. van de Kamp, P. A. *Rev. Astr. Astrophys.* **9**, 103–126 (1971).
19. Harrison, E. R. *Nature* **270**, 324–326 (1977).
20. Henrichs, H. F. & Staller, R. F. *Nature* **273**, 132–134 (1978).
21. Bignami, G. F., Caraveo, P. A. & Lamb, R. C. *Astrophys. J.* **272**, L9–L13 (1983).
22. Luyten, W. J. *NLT Catalogue*, Univ. Minnesota (1976).
23. Bahcall, J. N. & Soneira, R. *Astrophys. J.* **246**, 122 (1981).
24. Alvarez, W. & Muller, R. A. *Lawrence Berkeley Lab. Preprint LBL-17300* (1984); *Nature* **308**, 718–720 (1984).
25. Keller, G., D'Hondt, S. & Vallier, T. *Science* **221**, 150–152 (1983).

Evidence from crater ages for periodic impacts on the Earth

Walter Alvarez* & Richard A. Muller†

* Department of Geology and Geophysics; † Department of Physics and Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, USA

Recent evidence has indicated that the impact of a comet or asteroid may have been responsible for mass extinction at the ends of both the Cretaceous¹ and the Eocene²⁻⁴. Quantitative analysis by Raup and Sepkoski⁵ showed that mass extinctions occur with a 26-Myr period, similar to the period seen in qualitative pelagic records by Fischer and Arthur⁶. To account for the possibility of periodic comet showers, Davis *et al.*⁷ proposed that such showers could be triggered by an unseen solar companion star as it passes through perihelion on a moderately eccentric orbit. To test a prediction implicit in this model we examined records of large impact craters on the Earth. We report here that most of the craters occur in a 28.4-Myr cycle. Within measurement errors, this period and its phase are the same as those found in the fossil mass extinctions. The probability that such agreement is accidental is 1 in 10³.

A recent compilation by Grieve⁸ lists 88 dated craters for which signs of shock metamorphism suggest probable impact origin. The known craters are located primarily in stable, well-studied regions in North America, Europe, Australia and the USSR. There are no known impact craters on the sea floor. The list shows a strong bias towards recent craters, most of which will probably be removed by erosion in the near future; 12 of the 88 dated craters have ages within the past 5 Myr.

We restricted ourselves to impact craters in roughly the range of ages in which Raup and Sepkoski saw the periodicity: 5–250 Myr BP. The lower limit of 5 Myr was chosen to reduce the bias from the large number of craters surviving from the recent past. In order to be able to see periods as short as 26 Myr, we included only craters whose age uncertainty is ≤ 20 Myr. Table 1 lists 13 craters that meet this criterion. There have been some improved measurements since Grieve's compilation, and some older ages need to be revised on the basis of new standardized decay constants. After a search of the literature we developed our own revised values for the crater ages and their uncertainties; these are listed in Table 1. To avoid any possibility of bias, however, we confined the bulk of our present analysis to the values given by Grieve rather than our own. We discuss later the effect on our analysis of using the revised numbers instead of Grieve's values.

In Fig. 1a a rectangle of unit area has been used to represent each crater having a diameter >10 km. (We will discuss the sensitivity of our analysis to the choice of diameter cutoff later.) The width of the rectangle represents twice the standard deviation error for the age, and overlapping rectangles have been stacked. In Fig. 1b the rectangles have been replaced by gaussian distributions, with the r.m.s. for each crater set by the age uncertainty; this plot represents our best statistical representation of the history of impacts on the Earth during the past 250 Myr. A periodicity of about the same frequency and phase as that found in the Raup and Sepkoski analysis is evident, as indicated by the arrows placed at ~ 28 -Myr intervals.

Figure 2 shows the Fourier power spectrum of Fig. 1b. (We found no essential difference between the Fourier transforms of Fig. 1a and b.) In this plot the peak near the frequency 0.035 Myr⁻¹ corresponds to a period of 28.4 Myr, with the first maximum at 13 Myr. The arrows in Fig. 1 actually correspond to the frequency and phase found in this Fourier transform. Below we will reconcile our 28.4-Myr period with the 26-Myr period in the palaeontological extinctions.

To understand the statistical significance of the peak in the Fourier power spectrum, we generated Monte Carlo simulations

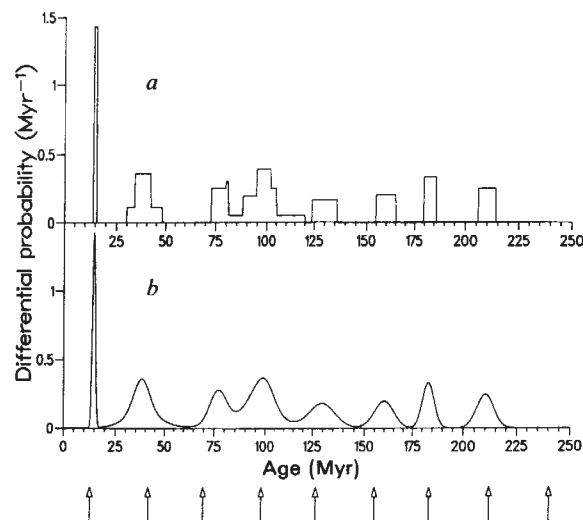


Fig. 1 Impact craters on the Earth having a diameter >10 km and age between 5 and 250 Myr. Only craters having age uncertainties of ≤ 20 Myr as listed by Grieve⁸ have been used. In a, each crater is represented by a rectangle of unit area and width equal to twice the age error. In b, each crater is represented by the equivalent gaussian curve. The arrows indicate the frequency and phase found from the best fit to the data.

of craters having random ages; 1,000 sets were generated, each containing 11 craters with random ages between 5 and 250 Myr, but with age errors (that is, gaussian widths) identical to and of the same order as those from the real craters. The Fourier transform was calculated for each of the simulated sets; the average of the 1,000 power spectra is indicated by a dotted line in Fig. 2. In each of the 1,000 Fourier power spectra we searched for peaks as high as that seen in the real data; such peaks occurred for frequencies equal to or above 0.035 Myr⁻¹ in 8 of the 1,000 randomly generated sets. From this analysis we conclude that the confidence level for the existence of the periodicity is $\sim 99\%$. Other analyses (for example, finding the number of large peaks in the 1,000 Monte Carlo sets having periods within 2 Myr of the 26-Myr period of Raup and Sepkoski; analysis of the statistics of the peaks; χ^2 calculations; simulations with craters weighted according to diameter) gave confidence levels in the range 97–99.5%. If we take into account the agreement of our periodicity not only in frequency but also in phase with the extinction events of Raup and Sepkoski, we calculate that our confidence level is closer to 99.9%; that is, the probability of random craters giving a signal as large as the one we see, agreeing within 2 Myr in period and 2 Myr in phase.

Figure 3 shows another way of displaying the periodicity (suggested by S. Perlmutter): for every pair of craters in our list of 11 large craters, the difference in their ages was plotted as a gaussian with the two errors combined in quadrature; these gaussians were then superimposed. Peaks are evident in this plot for differences of 28.4 Myr and all multiples thereof.

To estimate the accuracy of our frequency and phase determination, we performed a Monte Carlo generation of 20 sets of craters in each of which the 11 craters were initially forced to occur at precisely 28-Myr intervals, and then randomly 'jittered' (using gaussian statistics) according to the uncertainties in the real crater data. The Fourier transform was then taken, and a best-fit frequency and phase determined. From these simulations we estimate the uncertainty in our period to be ± 1 Myr, and the uncertainty in the time of the first maximum to be ± 2 Myr.

The age of the first crater maximum, 13 ± 2 Myr, is identical with the age of the first 'extinction event' of Raup and Sepkoski. One might worry that our slightly different period (28.4 ± 1 Myr as opposed to 26 Myr) would cause our cycles and theirs to be

Table 1 Impact craters of ≥ 5 km diameter having ages between 5 and 250 Myr with 1 s.d. age uncertainties of ≤ 20 Myr

Crater no.	Diameter (km)	Age (Myr)	Revised age†	Ref.	Location
43	10		7 ± 4	15	Karla, USSR
35	20		13 ± 11	10	Haughton, Canada
73 (88*)	24	14.8 ± 0.7			Ries, Germany
60	28	38 ± 4			Mistastin, Labrador
99	8.5	37 ± 2			Wanapitei, Ontario
69	100	39 ± 9			Popigai, Siberia
50	14	77 ± 4	78 ± 2	11	Lappajarvi, Finland
87	25	95 ± 7			Steen River, Alberta
18	25	100 ± 5			Boltys, Ukraine
52	17	100 ± 20			Logoisk, USSR
56	5	118 ± 2	119 ± 2	14	Mien Lake, Sweden
33	22	130 ± 6	133 ± 6		Gosses Bluff, Australia
74	23	160 ± 5			Rochechouart, France
65	15		185 ± 10	12	Obolon, Ukraine
70	80	183 ± 3			Puchezh-Katunki, USSR
54	70	210 ± 4	214 ± 3	13	Manicouagan, Quebec

The first age given is taken from Grieve⁸, and it is the primary one used in the present analysis. An age value was only included if there was an error estimate available. Only those craters in this list that have a diameter of ≥ 10 km are used in Figs 1–3.

* As these two craters are probably from the same event, we included only the larger one.

† Our own estimates based on current values for decay constants, and on radiometric and palaeontological data in the references cited.

out of phase after 100 Myr. However, there are uncertainties in the age determination of the palaeontological boundaries that increase significantly for ages > 100 Myr. Figure 4 plots a band representing our 28.4 ± 1 Myr period together with the best estimates for the palaeontological ages and errors, as evaluated by Harland *et al.*⁹. With the possible exception of the Tithonian event, a 28.4-Myr periodicity is a good fit to the extinction events of Raup and Sepkoski. Note that there is a slippage of one cycle between the crater events and the extinction events, so that the Permian–Triassic event is cycle 10 in the extinction sequence and cycle 9 in the crater sequence. The cycle slippage occurs in the 150–200 Myr interval, when there are three minor extinction events (7–15% of the families dying out) but only two predicted. We conclude that the record is not well defined in the region of slippage. This is supported by the fact that when Raup and Sepkoski divided the extinction sequence into two halves, they found that 27–29-Myr periodicities were significant to better than 95% confidence levels in both halves. Fischer and Arthur's qualitative cyclicity record⁶ shows one less cycle than ours, with the Permian–Triassic event at cycle 8; again the cycle slippage occurs in the 150–200-Myr region, which is clearly the weakest part of the palaeontological–stratigraphical record. The ages of

the extinction events are best determined in the most recent four cycles; this is the period for which we test the agreement in phase between the extinction events and the crater ages.

To test the sensitivity to our chosen cutoffs in age and crater size, we have repeated much of the above analysis for different values of these parameters. We varied the crater diameter cutoff continuously from 0 to 20 km; the 28.4-Myr peak was significant at all these values, and reached its maximum intensity for a 5 km cutoff (20% higher than the peak in Fig. 2). Inclusion of the data below 5 Myr or above 250 Myr did not significantly affect the peak height, but it did increase the background level. We also tried removing craters from our list to test the sensitivity of our analysis to the presence or absence of particular craters. The only crater in our nominal list of 11 which does not contribute to the 28.4-Myr periodicity is Lappajarvi, 77 ± 4 Myr old. When we lowered the diameter threshold, adding craters, we found a second peak in the power spectrum at 0.047 Myr^{-1} (period of 21 Myr). A peak at this frequency is also present in Fig. 2 but with less apparent significance. As our noise estimates had been made for random distributions of craters, to investigate the meaning of this peak we generated Monte–Carlo data sets which had a sequence of craters with a real 21-Myr or 28-Myr

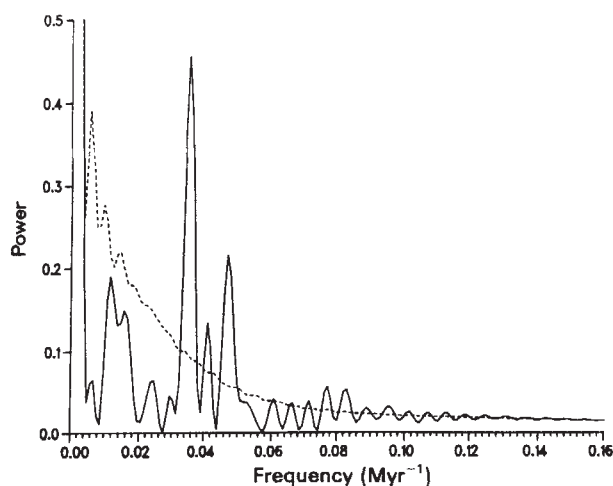


Fig. 2 Fourier power spectrum of Fig. 1b. The large peak at 0.035 Myr^{-1} corresponds to a period of 28.4 Myr. The dotted line is a background estimate, taken from the average of 1,000 Monte Carlo-generated data sets, each with a random distribution of crater ages but having the same age uncertainties as the real data.

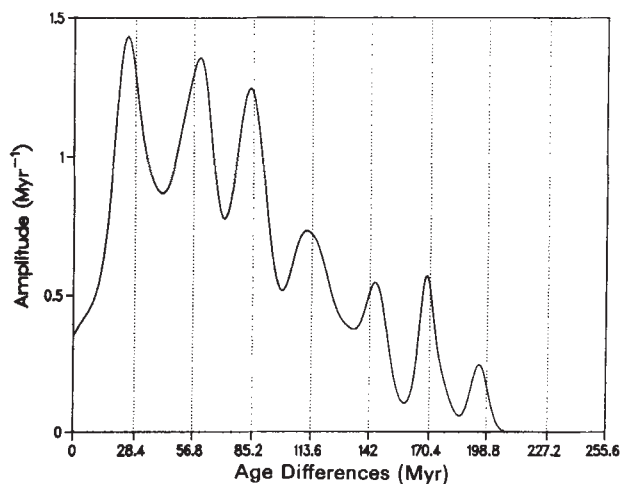


Fig. 3 Distribution of age differences in the crater data. For every possible pair chosen from the 11 craters used in the analysis, the difference in ages was plotted as a gaussian with width equal to the quadratically combined errors for the two ages. The gaussians for all the pairs were then added. The vertical lines are plotted at nominal age differences of 28.4 Myr and multiples thereof.

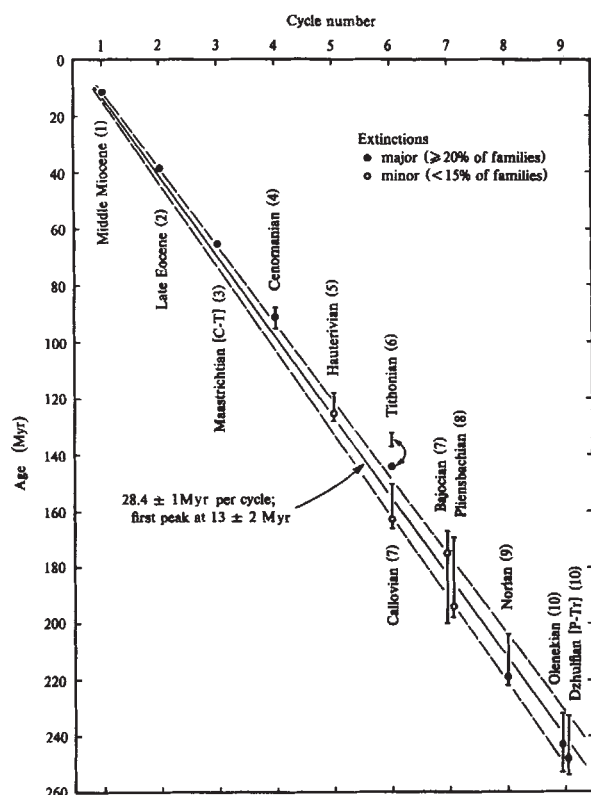


Fig. 4 Comparison of the periodicity found in the crater data with the extinction events from the fossil record. The band represents the periodicity found in the Fourier analysis of the crater data. The data points and confidence limits refer to the 'extinction events' of Raup and Sepkoski⁵. The two largest extinctions occur at the Cretaceous-Tertiary (C-T) and Permian-Triassic (P-Tr) boundaries. The error bars assigned to these events are derived from the review of the time scale by Harland *et al.*⁹. Cycle numbers for extinction events are given in parentheses; slippage of one cycle between the two data sets is discussed in the text. Because of the constraint that palaeontological stages must occur in the proper sequence, model ages⁹ sometimes fall outside the range of relevant radiometric data, as in the case of the Tithonian.

periodicity, jittered by the age errors, in addition to a few craters with random ages. We found that in the presence of a real 28-Myr period, a spurious second peak often appeared at 21 Myr, but that in the presence of a real 21-Myr period there was no false 28-Myr peak. Based on these simulations, we believe that there is no significant evidence for a real 21-Myr periodicity. The 21-Myr peak was strongly suppressed when we either increased the crater diameter threshold above 2 km, or weighted the contribution of the craters to the Fourier transform by a factor proportional to their diameters.

As mentioned above, to avoid any possibility of bias in crater selection, we deliberately chose not to reevaluate any of the data in performing our initial analysis, but accepted the compilation of Grieve at face value. We now turn our attention to the revised ages¹⁰⁻¹⁵ in Table 1, which are based on our own extensive search of the literature. Palaeontological dates on lake beds within craters have allowed us to add three new craters to the list: Grieve 43 (10 km diameter, age 7 ± 4 Myr), 35 (20 km diameter, age 13 ± 11 Myr) and 65 (15 km diameter, age 185 ± 10 Myr). The latter two agree reasonably well with our periodicity, which includes 13 Myr and 183 Myr among its cycles. We also list four craters for which there are minor adjustments in the age values or their uncertainties; Grieve 50, 56, 33 and 54. These adjustments make no significant difference to our analysis,

except for Grieve 50; this is the 14 km crater at Lappajarvi which, as we noted, does not contribute to the spectrum at 28.4 Myr. With the age estimate tightened from 77 ± 4 to 78 ± 2 Myr, we can definitely exclude it from any cyclic process that has a narrow phase of activity, such as that proposed in the solar companion model⁷. It is vital to obtain smaller errors for the crater ages.

The period and phase of the crater data show strong agreement with those of the palaeontological extinctions, in support of the hypothesis that the extinctions were caused by periodic impact showers. The fact that we find typically one or more craters to associate with each cycle, despite the fact that Grieve's compilation covers less than 10% of the Earth's surface, implies that many impacts occur during each shower. This is corroborated by the discovery of at least two different levels of microtektites near the Eocene-Oligocene boundary¹⁶. However, given the relatively large crater-age uncertainties, we have, as yet, no indication that the showers have a duration as short as 1-2 Myr. The fact that the periodicity is found primarily among the larger craters suggests that there may be a random background of low-energy impacts in addition to the showers. This is plausible if the shower craters come from comets and the background craters from asteroids. Such background impacts, including an occasional very large one, are inevitable in view of the statistics of modern Apollo asteroids. The long-term cratering rate is considered to be compatible with the current flux of Apollo objects and comets¹⁷⁻²⁰. The periodicity of the crater ages may require a revision of this view. Comet velocities are higher than those of the asteroids, and half of the comets (but none of the asteroids) enter the Solar System with retrograde motion leading to head-on impacts. These two factors can increase the impact energies of comets over those of asteroids (assuming equal masses) by an order of magnitude. A modest improvement in the ages measured for a few craters should clarify the nature of the showers.

After this manuscript was submitted, we learned that a solar-companion hypothesis generally similar to that of Davis *et al.*⁷ had been independently proposed by Whitmire and Jackson (accompanying paper)²¹.

We thank Saul Perlmutter for his help with the analysis. We have benefited greatly from the critical analysis and suggestions of many of our colleagues, particularly L. W. Alvarez, F. Crawford and T. Mast. During this research, W.A. was supported by a fellowship from the John Simon Guggenheim Memorial Foundation, and R.A.M. by a fellowship from the John D. and Catherine T. MacArthur Foundation. The work was partially funded by the NSF Alan T. Waterman Award, by NSF grant EAR-81-15858, and by the Department of Energy under contract DE-AC0376SF00098.

Received 30 January; accepted 8 March 1984.

1. Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. *Science* **208**, 1095-1108 (1980).
2. Asaro, F., Alvarez, L. W., Alvarez, W. & Michel, H. V. *Snowbird Conf. Abstr., Lunar planet. Inst. Contr.* **449**, 2 (1981).
3. Ganapathy, R. *Science* **216**, 885-886 (1982).
4. Alvarez, W., Asaro, F., Michel, H. V. & Alvarez, L. W. *Science* **226**, 886-888 (1982).
5. Raup, D. M. & Sepkoski, J. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 801-805 (1984).
6. Fischer, A. G. & Arthur, M. A. *Soc. Econ. Paleont. Miner. Spec. Publ.* **25**, 19-50 (1977).
7. Davis, M., Hut, P. & Muller, R. A. *Lawrence Berkeley Lab. Preprint LBL-17298* (December 1983); *Nature* **308**, 715-717 (1984).
8. Grieve, R. A. F. *Geol. Soc. Am. Spec. Pap.* **190**, 25-37 (1982).
9. Harland, W. B. *et al. A Geologic Time Scale*, 131 (Cambridge University Press, 1982).
10. Frisch, T. & Thorsteinsson, R. *Arctic* **31**, 108-124 (1978).
11. Jessberger, E. K. & Reimold, W. U. *J. Geophys.* **48**, 57-59 (1980).
12. Val'ter, A. A., Gurov, Ye. P. & Ryabenko, V. A. *Dokl. Akad. Nauk SSSR Earth Sci. Sect. (transl.)* **232**, 37-40 (1977).
13. Jahn, B. M., Floran, R. J. & Simonds, C. H. *J. geophys. Res.* **83**, 2799-2803 (1978).
14. Bottomley, R. J., York, D. & Grieve, R. A. F. *Contr. Miner. Petrol.* **68**, 79-84 (1978).
15. Masaitis, V. L., Danilin, A. N., Darpo, G. M. & Raykhlin, A. I. *Dokl. Akad. Nauk SSSR Earth Sci. Sect. (transl.)* **230**, 48-51 (1976).
16. Keller, G., D'Hondt, S. & Vallier, T. *Science* **221**, 150-152 (1983).
17. Wetherill, G. W. & Shoemaker, E. M. *Geol. Soc. Am. Spec. Pap.* **190**, 1-13 (1982).
18. Weissman, P. R. *Geol. Soc. Am. Spec. Pap.* **190**, 15-24 (1982).
19. Shoemaker, E. M. A. *Rev. Earth-planet. Sci.* **11**, 461-494 (1983).
20. Shoemaker, E. M. *Proc. Dahlem Workshop, Berlin, May 1983* (Springer, Berlin, in the press).
21. Whitmire, D. & Jackson, A. *Nature* **308**, 713-715 (1984).

Fractal character of fracture surfaces of metals

Benoit B. Mandelbrot*, Dann E. Passoja†
& Alvin J. Paullay‡

* IBM Research Center, Yorktown Heights, New York 10598, USA

† Union Carbide Corporation, Tarrytown, New York 10591, USA

‡ Columbia University New York, Bronx Community College,
Bronx, New York 10753, USA

When a piece of metal is fractured either by tensile or impact loading (pulling or hitting), the fracture surface that is formed is rough and irregular. Its shape is affected by the metal's microstructure (such as grains, inclusions and precipitates, whose characteristic length is large relative to the atomic scale), as well as by 'macrostructural' influences (such as the size, the shape of the specimen, and the notch from which the fracture begins). However, repeated observation at various magnifications also reveals a variety of additional structures that fall between the 'micro' and the 'macro' and have not yet been described satisfactorily in a systematic manner. The experiments reported here reveal the existence of broad and clearly distinct zone of intermediate scales in which the structure is modelled very well by a fractal surface. A new method, slit island analysis, is introduced to estimate the basic quantity called the fractal dimension, D . The estimate is shown to agree with the value obtained by fracture profile analysis, a spectral method. Finally, D is shown to be a measure of toughness in metals.

Fractals are the concern of a new geometry¹, whose primary object is to describe the great variety of natural structures that are irregular, rough or fragmented, having irregularities of various sizes that bear a special 'scaling' relationship to one another. In very loose terms, they seem to fall into a regular hierarchy in which each level is an up-sized or down-sized version of the levels below or above it. Fractal geometry characterizes the scaling structure of a surface by a number D , called the fractal dimension, that can range from 2, when the surface is smooth, up to 3. The fractal dimensional increment of a surface is $D-2$. As it increases from 0 to 1, the irregularities become increasingly predominant and the notion of overall shape of the surface becomes progressively less meaningful. Similarly, the fractal dimension of a curve is at least 1, the fractal dimensional increment being $D-1$.

The term 'fractal' was chosen in explicit cognizance of the fact that the irregularities found in fractal sets are often strikingly reminiscent of fracture surfaces in metals (though not, for example in glass). However, metal fractures are only extremely crinkly (down to the limits of their microstructural size range), while fractals are infinitely crinkly. Hence, it is not possible to say that metal fracture is strictly a fractal. Nevertheless, metals resemble a fractal so closely that it makes good sense to use a metal for modelling a fractal. Our experiments in metal fracture show that D is very well defined for all the specimens examined, and takes on the same value for different specimens of the same metal having similar thermomechanical treatments.

Recent studies^{2,3} conclude that a metal fracture's spectrum is indicative of a fractal structure. A recent review⁴ also indicates that fractals might model metal fracture surfaces, and reports exploratory studies. The more refined tools that we use to analyse metal fracture surfaces make a stronger case and add to the existing tools of fractographic analysis^{4,5}.

Our first series of tests is exemplified by Fig. 1. A fractured steel specimen was plated with electroless nickel and mounted in an epoxy mount by vacuum impregnation in order to ensure edge retention. The specimen was then polished parallel to the plane of fracture. 'Islands' of steel surrounded by nickel appeared which, on subsequent polishing, grew and merged. We propose to call these structures slit islands. The islands

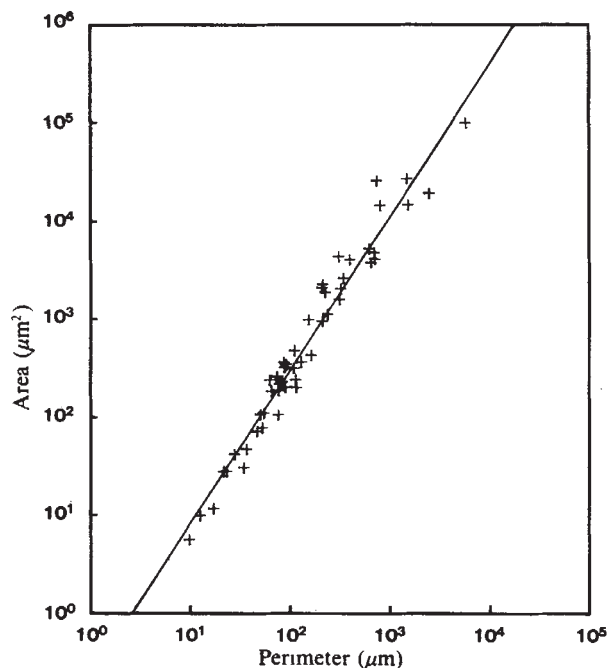


Fig. 1 Fractal area-perimeter relationship for slit islands. 300-Grade Maraging steel. Ruler = 1.5625 μm ; fractal dimensional increment = 1.28.

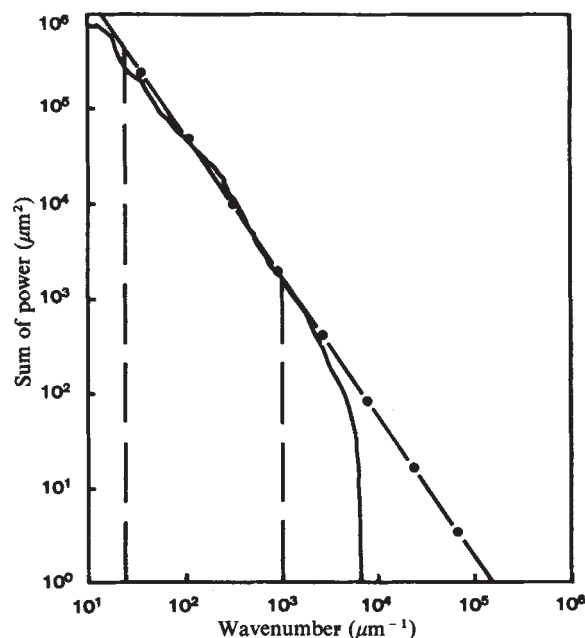


Fig. 2 Cumulative spectrum for vertical cuts. 300-Grade Maraging steel. Slope = 1.4878; fractal dimensional increment = 1.26. Average of five profiles.

contain 'lakes within islands' and 'islands within lakes'; we include the former and neglect the latter.

The slit island 'coastlines', being curves, are much easier to investigate than the surfaces themselves. Yet we have found that the perimeter-area relationship of fractal dimensional analysis (ref. 1, Chap. 12), if applied to these coastlines, reveals many essential facts about the surface. We propose that the resulting method of analysis, which is entirely new and very powerful, be called slit island analysis. When islands are derived from an initial fractal surface of dimension D by sectioning with a plane, their coastlines are of fractal dimension $D' = D - 1$. Thus, the fractal dimensional increments $D-2$ and $D'-1$ are equal. The theory of fractals suggests that the areas and

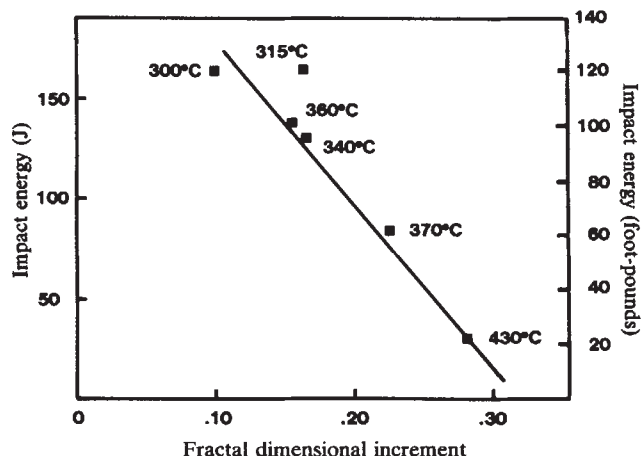


Fig. 3 Impact energy versus fractal dimensional increment. 300-Grade Maraging steel, room temperature, Charpy impact.

perimeters of these islands should be measured in the same way, and that one should trace the graph of $\log(\text{perimeter})$ versus $\log(\text{area})$. In a fractal, this graph is rectilinear of slope D' . Figure 1 shows that such is indeed the case for fracture surfaces.

Our second series of tests of fractal structure which complements the first, uses fracture profile analysis based on Fourier analysis³ (Fig. 2). Here, the plated and processed fracture was sectioned perpendicular to the fracture surface to expose it in profile. As usual, the profile's sample spectra exhibit wide oscillations. Some are statistical artefacts, but others reflect fundamental lengths of the microstructure and their accompanying higher order harmonics. Statistical fluctuations must be averaged out and the fluctuations due to the microstructure should be prevented from overly affecting the analysis. To achieve both goals, we averaged five spectra taken from serial sections, and integrated them from high frequency to low. Now, the fractal hypothesis requires the integrated spectrum to take the form $k^{-B'}$, with $B' = B - 1 = 6 - 2D$. The fractal character of the surface was tested by plotting the data on doubly logarithmic coordinates to see if the curve has a large straight central portion of absolute slope B' . The plots (Fig. 2) are indeed straight but over narrower ranges than the area-perimeter plots. Thus, the fracture surface is again inferred to be fractal, and its fractal dimension is estimated by $D = 3 - B'/2$.

When a method of analysis yields a nearly straight graph, it is tempting to measure the 'local' slope of this curve. The goal would be to identify the harmonics of the microstructure and other features that (contrary to the overall fractal behaviour) are defined by well-defined length scales. However, as the local slope is severely affected by statistical fluctuations, attempts to interpret its variations are useless. Similarly, the 'spectrum of fractal dimension' described in ref. 5 mostly reflects the underlying noise.

Careful analysis of some of our data indicates that the central region of the log-log plots splits into two distinct subregions characterized by different values for D . Such a juxtaposition of different fractal zones is very familiar in other applications of fractals. The scales of the clear-cut cross-over points between zones tend to be significant with respect to structure and deserve further investigation. For other metals, the diagrams are even more complex.

To relate D to known metallurgical quantities, we took a series of identical 300 grade Maraging steel Charpy impact specimens, and heat-treated them at different temperatures. Figure 3 shows that the value of D decreases smoothly with an increase of the impact energy as measured by a standard Charpy impact test. This relationship must reflect the changes in the microstructure that occur during ageing.

In our view, a fracture that is transgranular (that is, runs through the grains) involves an atypical form of the notion of percolation. For example, during fracture of a ductile material,

any voids forming around inclusions increase in size and coalesce into void sheets; these ultimately form the fracture surface. If the growth of a void were independent of its neighbours or its position in the specimen, we would be faced with a process physicists call percolation. Its applicability here is associated with load sharing among the tensile ligaments that remain; void sheet coalescence is merely percolation of a crack in the microstructure leading to final fracture. Percolation would imply that the fracture's fractal dimension takes some universal value independent of the material, and dependent only on the properties of space. But percolation is a very crude model here. As soon as the initial void growth has coalesced locally into small sheets, the strains on the supporting ligaments increase, and surrounding voids grow at a rate that varies with their position in the specimen. Certainly, spatial variability associated with the microstructure is dependent on the microstructure and the resulting process differs from the usual percolation. Were this view to be confirmed by further experimentation, the study of fracture could benefit from techniques used in the study of phase transitions in statistical physics.

By showing linear relationships among all available data while covering such a broad range of sizes, Figs 1 and 2 are almost unique in metallurgy. Their linearity demonstrates that the fracture surfaces present in several grades of steel produced by impact or uniaxial tensile loading are fractal.

The metallographic assistance of H. I. Kohlmorgen is gratefully acknowledged. Comments by J. A. Given helped clarify this text.

Received 4 November 1983; accepted 18 January 1984.

1. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, New York, 1982); also *Les objets fractals* (Flammarion, Paris, 1975, 1984) and *Fractals: Form, Chance and Dimension* (Freeman, New York, 1977).
2. Passoja, D. E. & Psioda, J. A. in *Fractography and Materials Science* (eds Gilbertson, L. N. & Zipp, R. D.) 335-386 (American Society for Testing Materials, Philadelphia, 1981).
3. Passoja, D. E. & Amborski, D. J. in *Microstruct. Sci.* 6, 143-148 (1978).
4. Gilbertson, L. N. & Zipp, R. D. (eds) *Fractography and Materials Science*, ASTM STP 733 (American Society for Testing Materials, Philadelphia, 1981).
5. Coster, M. & Chermant, J. L. *Int. Metals Rev.* 28, 234-238 (1983).

Atmospheric transport of pollutants from North America to the North Atlantic Ocean

R. C. Harriss, E. V. Browell, D. I. Sebacher, G. L. Gregory, R. R. Hinton, S. M. Beck, D. S. McDougal & S. T. Shipley

NASA Langley Research Center, Hampton, Virginia 23665, USA

Ground-based measurements strongly support the hypothesis that pollutant materials of anthropogenic origin are being transported over long distances in the mid-troposphere and are a significant source of acid rain, acid snow, trace metal deposition, ozone and visibility-reducing aerosols in remote oceanic and polar regions of the Northern Hemisphere¹⁻⁴. Atmospheric sulphur budget calculations⁵ and studies of acid rain on Bermuda⁶ indicate that a large fraction of pollutant materials emitted into the atmosphere in eastern North America are advected eastwards over the North Atlantic Ocean. We report here the first direct airborne measurements of the vertical distribution of tropospheric aerosols over the western North Atlantic. A newly developed airborne differential adsorption lidar (DIAL) system⁷ was used to obtain continuous, remotely sensed aerosol distributions along its flight path. Our data document two episodes of long-distance transport of pollutant materials from North America over the North Atlantic Ocean.

Documenting source-transport-receptor relationships to pollution studies is an elusive, complex problem which con-

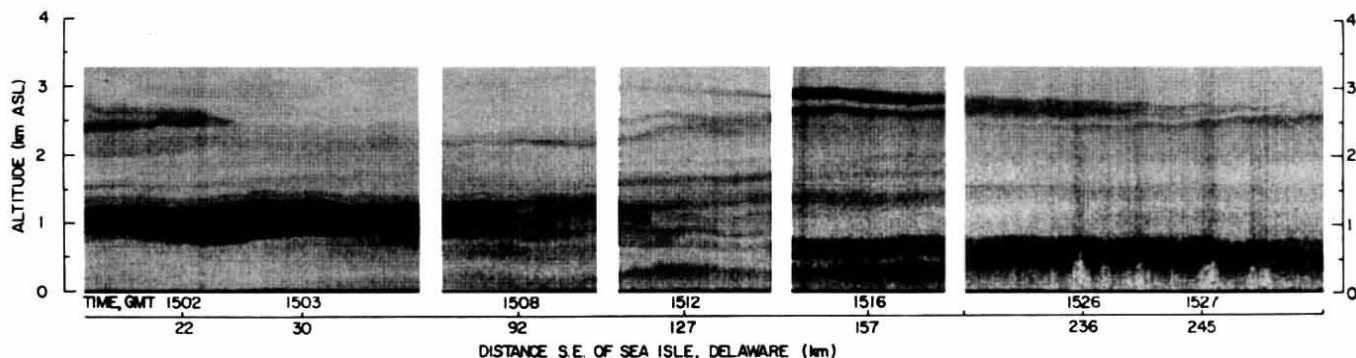


Fig. 1 Selected lidar aerosol cross-sections taken along a flight line 15–250 km south-east of Sea Isle, Delaware, August 1982.

tributes to most of the controversy over the origin of acid deposition in non-urban areas. While the combination of ground-based chemical measurements, meteorological trajectory analyses and qualitative satellite observations have contributed to important advances in long-range transport studies, direct measurements of tropospheric photochemical processes, transport dynamics and deposition are required to resolve many scientific and policy issues. Groundbased and airborne active remote sensing techniques offer promise for such critical direct measurements of mesoscale spatial and temporal variations of aerosols and gases in the troposphere^{7,8}. To test the ability of an airborne lidar system to measure long-range transport, we conducted preliminary experiments over the western North Atlantic in July 1981 and August 1982, in the region between Wallops Island, Virginia, and Bermuda. The NASA Langley DIAL system transmits a laser pulse energy of 80 mJ through a quartz window in the bottom of the aircraft. Aerosol profiles along the lidar line of sight are obtained in real time on the aircraft from an analysis of backscattered laser light at a wavelength of 600 nm. The spatial resolution for aerosol measurements is 15 m in the vertical and 20 m in the horizontal. A complete discussion of the DIAL system and its measurement capabilities is available in ref. 7.

In situ measurements of O₃, CO, air temperature, dew point, wind speed, wind direction, aerosol size and aerosol composition were made along flight transects consisting of 45 min at specific altitudes selected in-flight on the basis of atmospheric aerosol features observed in the lidar aerosol profiles. *In situ* O₃ data were collected continuously by a chemiluminescent airborne system with an absolute accuracy of $\pm 10\%$ and precision of $\pm 3\%$ (ref. 9). Grab samples for CO determination were collected in stainless steel bottles previously heated to 1,000 °C to remove contaminants. These samples were analysed by flame

ionization gas chromatography¹⁰ immediately after flights. Accuracy and precision of the CO measurements are both $\pm 2\%$ based on a 1.08 parts per million (p.p.m.) standard reference gas. Aerosols were collected on Teflon-coated quartz fibre filters and analysed in the laboratory by ion chromatography.

Previous studies have shown that summertime haze events can result from complex mixing of air masses over the entire eastern United States¹¹. During both field experiments reported here, the atmosphere was cloud-free along the flight line, with winds at all altitudes sampled (1–5 km) flowing from west to east at 2–8 m s⁻¹. The GOES satellite images for the period 4 and 5 August 1982 indicate an extensive source region of atmospheric haze over the eastern United States with a plume extending ~ 300 km over the western North Atlantic. Meteorological back-trajectory analysis for the 850- and 700-mbar levels also indicates that winds in the sector studied flowed from west to east during our July 1981 measurements. Back-trajectories for 4 and 5 August 1982 indicate a south-west to north-east flow.

Selected lidar aerosol cross-sections are illustrated in Figs. 1–3. Relative variations in aerosol backscatter cross-sections are illustrated with a 16-tone grey scale; darker shading indicates higher backscatter, which can be related to higher aerosol concentrations. Features observed during both experiments include: (1) A major concentration of aerosol at altitudes of 650–1,800 m above sea level (a.s.l.). This feature was continuous over the flight lines from the coastline to distances up to 250 km offshore. (2) A marine boundary layer enrichment of aerosol from the sea surface to approximately 250–300 m a.s.l. During the August 1982 experiment, up to seven distinct layers of aerosol enhancement were observed at intervals along the flight line between the sea surface and 3,000 m a.s.l. Such layers maintained distinct vertical integrity over hundreds of kilometres, generally becoming less intense in an eastward direction.

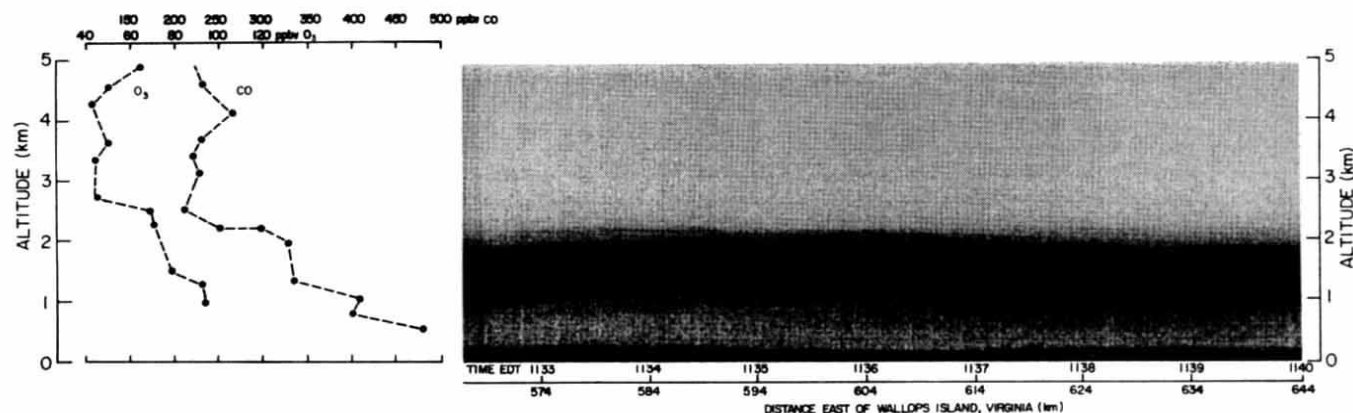


Fig. 2 Continuous lidar aerosol cross-section taken along a flight line 564–644 km east of Wallops Island, Virginia, July 1981. Average concentrations of O₃ and CO collected with *in situ* methods are compared with the remotely sensed aerosol distribution; p.p.b.v., parts per 10⁹ by volume.

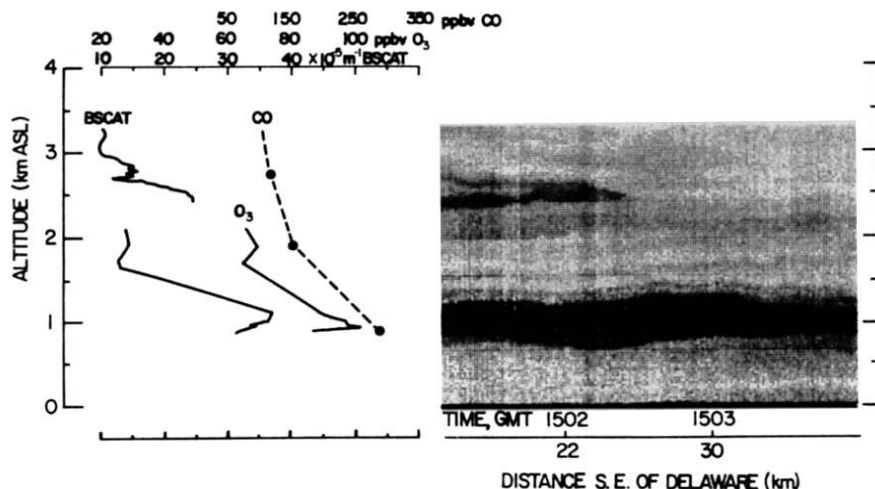


Fig. 3 Selected lidar aerosol cross-sections taken along a flight line 230–260 km south-east of Sea Isle, Delaware, August 1982. Averaged *in situ* measurements of O_3 , CO and BSCAT (aerosol light scattering) are compared with the remotely sensed aerosol distribution. Solid lines indicate continuous *in situ* data, dashed lines indicate no data between selected sampling altitudes.

Figures 2 and 3 summarize *in situ* data on the O_3 , CO and scattering properties of marine boundary layer (< 300 m a.s.l.), aerosol-rich intermediate altitude air (300–3,000 m a.s.l.) and relatively clean upper tropospheric air (> 3,000 m a.s.l.). During both experiments, minimum concentrations of all gaseous and particulate species were observed in the upper troposphere. Aerosol-rich layers being advected off the east coast of the United States at approximately 1 km a.s.l. had the highest concentrations of CO, O_3 and total particulates. A comparison of 9 July 1981 vertical profiles of aerosols, O_3 and CO in the region approximately 400–600 km east of the United States coastline shows a positive correlation below 2 km a.s.l. and an anticorrelation between O_3 and CO in the low aerosol, mid-tropospheric region from 2 to 5 km a.s.l. (Fig. 2). These data support the hypothesis of Fishman and Crutzen¹² that *in situ*

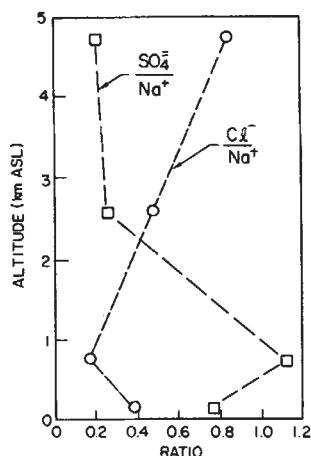


Fig. 4 Sulphate-to-sodium and chloride-to-sodium ratios in bulk aerosols collected at four altitudes along a flight line 10–390 km south-east of Sea Isle, Delaware. The sea water ratio is $SO_4/Na = 0.24$ and $Cl/Na = 1.8$.

production of tropospheric O_3 occurs in response to surface emissions of precursors. During the period of these measurements, a persistent inversion was present over the entire region, trapping surface emissions in the atmospheric boundary layer. During flights on 5 August 1982, layers of air with relatively high aerosol, CO and O_3 were present up to altitudes of 3 km a.s.l., with a pronounced enhancement of all three parameters at 1 km a.s.l. (Fig. 3). The water-soluble materials in the dominant aerosol layer of August 1982 are enriched in sulphate and depleted in chlorine relative to sodium (Fig. 4), indicating

possible gas-particle reactions between gaseous sulphur and nitrogen species with sea-salt aerosols. These observations support previous ground-based studies which hypothesized that gaseous H_2SO_4 and HNO_3 react with sea-salt resulting in the conversion of particulate chlorine into gaseous HCl (ref. 13).

Several results of these first continuous two-dimensional lidar observations of aerosol distribution from the coast of the United States over the western North Atlantic are important to future studies in tropospheric chemistry. First, aerosol and trace gas layers being advected off the United States in a stable west to east flow regime can maintain vertical integrity over distances of at least hundreds of kilometres. This vertical stratification implies that relatively thin layers of the atmospheric column can be very important to regional chemical budgets and cycles. Lidar aerosol tracers now provide the opportunity to make direct observations of atmospheric structure and transport phenomena. Such measurements can guide *in situ* sampling strategies for studies of photochemical processes during long-distance transport, with regard to particle conversion and deposition. The power of the lidar remote sensing technique is best illustrated by considering that during the mission from Wallops Island, Virginia, to Bermuda in August 1982, as reported here, we measured the equivalent of $\sim 10^5$ individual profiles of aerosol with a vertical resolution of 15 m. Using standard *in situ* techniques, no more than 10 profiles could have been obtained. Second, these preliminary measurements on the vertical distribution of O_3 , CO and aerosols over the western North Atlantic provide the first direct observation of how pollutants are being transported in both the boundary layer and free troposphere from source regions in the eastern United States to the North Atlantic.

We thank the flight crew of the NASA Electra and the staff of the Wallops Flight Center for their contribution to this research. C. Butler, A. Carter, L. Overbay, R. Allen, M. Mayo, J. Siviter and W. McCabe contributed to the UV DIAL development and operation. The US Naval Air Station, Bermuda, and NASA Tracking Station, Bermuda, provided valuable logistics support. This research was supported by the NASA Tropospheric Chemistry programme.

Received 2 September 1983; accepted 31 January 1984.

1. Rahn, K. A. & McCaffrey, R. J. *Nature* **280**, 479–480 (1979).
2. Kaerner, R. M. & Fisher, D. *Nature* **295**, 137–140 (1982).
3. Nielsen, T., Samuelsson, U., Grennfelt, P. & Thomsen, E. L. *Nature* **293**, 553–555 (1981).
4. Miller, J. M. & Yashinaga, A. M. *Geophys. Res. Lett.* **8**, 779–782 (1981).
5. Galloway, J. N. & Whelpdale, D. M. *Atmos. Environ.* **14**, 409–417 (1980).
6. Jickells, T., Knap, A., Church, T., Galloway, J. N. & Miller, J. *Nature* **297**, 55–57 (1982).
7. Brownell, E. V. *et al. Appl. Opt.* **22**, 522–534 (1983).
8. Pelon, J. & Megie, G. *Nature* **299**, 137–139 (1982).
9. Gregory, G. L., Hudgins, C. H. & Edahl, R. A. *Environ. Sci. Tech.* **17**, 100–103 (1983).
10. Swinnerton, J. W., Linneboom, V. I. & Cheek, C. H. *Limnol. Oceanogr.* **13**, 193–195 (1963).
11. Wolff, G. T., Kelly, N. A. & Ferman, M. A. *Science* **211**, 703–705 (1981).
12. Fishman, J. & Crutzen, P. J. *Nature* **274**, 855–858 (1978).
13. Martens, C. S. & Harriss, R. C. *J. geophys. Res.* **78**, 949–957 (1973).

Radiometric determination of the growth rate of *Nautilus* in nature

J. Kirk Cochran*

Department of Chemistry, Woods Hole Oceanographic Institution,
Woods Hole, Massachusetts 02543, USA

Neil H. Landman

Department of Invertebrates, American Museum of Natural History,
Central Park West at 79th Street, New York,
New York 10024, USA

Nautilus is the sole surviving genus of externally shelled cephalopods and therefore, determination of its rate of growth is important for understanding the life histories of a host of related fossil forms. Direct observations of its growth rate are difficult, however, so we have used here naturally occurring radionuclides. We used the $^{210}\text{Po}/^{210}\text{Pb}$ activity ratio as a chronometer to estimate the growth rates of seven specimens of *Nautilus belauensis* Saunders. In three immature animals the age of the last septum ranged from 10 to 52 days and, based on the age difference between the last two septa, the rate of septal formation was >120 to ~230 days. In four animals near or at maturity the age of the last septum was 110–240 days and the rate of septal formation 100–>290 days. Based on these rates the time to maturity is more than 10 yr. Rates of septal formation translate approximately into rates of <0.12 and <0.08 mm of ventral circumference per day for immature and mature animals, respectively, and agree well with mark-recapture observations of this species in Palau¹.

We have previously shown² that *Nautilus pompilius* incorporates the radioactive isotope ^{210}Pb ($t_{1/2}$ 22.3 yr) from seawater into its shell during growth. The granddaughter of ^{210}Pb , ^{210}Po ($t_{1/2}$ 138 days), is excluded during growth and its presence in the shell is a function of the radioactive decay of ^{210}Pb . Therefore, the $^{210}\text{Po}/^{210}\text{Pb}$ activity ratio furnishes a natural chronometer of growth. We analysed seven specimens of *N. belauensis* supplied by W. B. Saunders (Bryn Mawr) and captured as part of his tag-release programme in Palau¹, and although recapture data were not available for these seven specimens, their study provided the opportunity to compare radiometric results with field observation. The seven specimens were identified as immature to mature based on external shell characteristics (W. B. Saunders, personal communication, Table 1). Shells were cut in median section to facilitate sampling and measurement of septal number, thickness and angle. The shells ranged over 132–210 mm in diameter and 28–36 septa.

The four to six most recent septa of each specimen were removed and the radiochemical analyses performed according to methods already established². Septa were treated with H_2O_2 to eliminate organic matter and then rinsed briefly with 0.5 M HCl to remove any adsorbed ^{210}Po or ^{210}Pb which had been released and re-adsorbed during the H_2O_2 treatment. Samples of 0.05–6.0 g were dissolved in 6 M HCl and ^{208}Po tracer and stable Pb carrier were added. Po was plated onto silver disks and Pb was purified by ion exchange. The Pb yield was determined by atomic absorption analysis and ^{210}Pb was analysed by a second ^{210}Po plating after storage of the purified ^{210}Pb fraction.

The radiochemical data are presented in Table 2. In all samples the initial ^{210}Po separation (by plating) was completed within 26 days of collection. Uncertainty on the $^{210}\text{Po}/^{210}\text{Pb}$ activity ratio is $\sim\pm 10\%$. Samples for which the activity ratio at analysis was 1.0 ± 0.1 were not corrected to the time of collection. All other values have been corrected for ^{210}Po ingrowth from the time of collection to plating. Activity ratios include

the 1σ counting uncertainty in both ^{210}Po and ^{210}Pb measurements.

The $^{210}\text{Po}/^{210}\text{Pb}$ activity ratio of the most recent septum in all specimens was significantly less than 1.0. In immature animals 1, 2 and 7, radioactive disequilibrium persisted through the three or four most recently deposited septa. By contrast, in specimens 3, 5 and 6, near or at maturity, the activity ratio approached equilibrium by the second septum.

If no process other than radioactive decay affects the $^{210}\text{Po}/^{210}\text{Pb}$ activity ratio after formation of a septum, the activity ratio will change as a function of time according to:

$$A_{\text{Po}}/A_{\text{Pb}} = 1.0179 [1 - \exp(-\lambda_{\text{Po}} + \lambda_{\text{Pb}})t] \quad (1)$$

where A_{Po} is the activity of ^{210}Po , A_{Pb} is the activity of ^{210}Pb , λ_{Po} is the decay constant for ^{210}Po ($5.01 \times 10^{-3} \text{ day}^{-1}$), λ_{Pb} is the decay constant for ^{210}Pb ($8.51 \times 10^{-5} \text{ day}^{-1}$) and t is the average age of the septum (days). The septal age (t) calculated from equation (1) is a mean over the length of time of septum formation, and differences in t between adjacent, completely formed septa represent times between successive points in septal formation.

The most recent septum in each specimen was assigned an age based on the measured value of its activity ratio according to equation (1). Judging from age and septal thickness (Table 1), immature specimens 1, 2 and 7 were still in the process of forming their most recent septum, which had an age range of 10–52 days. The septum with an age of 10 days in specimen 7 was approximated but had a thickness only 4% that of the preceding septum. By contrast, the most recent septum in the four specimens near or at maturity (specimens 3–6) was thickened and ranged in age from 110 to 240 days. These data suggest that the rate of septal formation in immature animals is substantially greater than 30 days and in mature animals may exceed 240 days.

Another estimate of the time scale of septal deposition is provided by the age difference between adjacent septa (equation (1)). In the specimens analysed, the best resolution in age difference was obtained by consideration of the two most recent septa. Beyond this point, the $^{210}\text{Po}/^{210}\text{Pb}$ activity ratio approached equilibrium and the uncertainties associated with the measurement reduced the precision with which age differences could be resolved. All septa with $^{210}\text{Po}/^{210}\text{Pb}$ activity ratios ≥ 0.90 were assigned a minimum age of 400 days (about three half-lives of ^{210}Po). In specimens 3, 5 and 6 near or at maturity, radioactive equilibrium was reached by the second septum and therefore the age difference calculated between the two most recent septa provided a minimum estimate of the rate of septal formation. In immature specimens 1, 2 and 7, the calculated age difference also provided a minimum estimate of the rate of septal formation because the last septum was still incompletely formed. The error in the age difference is $\pm 20\%$ except for specimen 4 ($\pm 100\%$) and was calculated using the counting uncertainties on the activity ratios to set upper and lower limits.

The rates of septal formation (Table 1) ranged from >120 to ~230 days in immature specimens 1, 2 and 7 to 100 to >290 days in specimens 3, 4, 5 and 6 near or at maturity. These rates are long and exceed most estimates derived from aquarium studies of *Nautilus macromphalus* and *N. pompilius* which range over 30–130 days^{3–5}. They also exceed estimates of 75 and 25 days derived using the same radiometric technique on two *N. pompilius* (although the lower of these two values may be too low because of the larger uncertainty associated with a longer time between collection and analysis in that specimen)². Our data agree well, however, with Saunders' mark-recapture observations¹, in which two animals, immature at release, had secreted an additional septum by the time of recapture after 342 and 353 days.

The rates of septal formation in *N. belauensis* can be used to estimate the time to maturity. The first seven septa are deposited embryonically² and are therefore neglected in the calculations. In a hypothetical individual with 36 septa, if we assume an immature rate of growth of at least 120 days per septum to

*Present address: Marine Sciences Research Center, State University of New York at Stony Brook, Stony Brook, New York 11794, USA.

Table 1 Morphological and growth rate data for specimens of *Nautilus belauensis* arranged by stage of development at time of capture

Specimen	Stage of development at capture	No. of septa	Diameter (mm)	Approximation*	Thickness of septum 1†/ thickness of septum 2	Age of septum 1 (days)	Rate of septal formation‡ (days)	deg per day§	mm per day
WH1 (2733)	Immature	28	132	No	1.00	28	~ 230	~ 0.10	~ 0.10
WH2 (2830)	Immature	31	152	No	0.66	52	> 168	< 0.11	< 0.13
WH7 (2832)	Immature	36	186	Yes	0.04	10	> 120	< 0.16	< 0.14
WH5 (2739)	Immature, close to maturity	33	210	No	1.28	110	> 290	< 0.06	< 0.10
WH4 (2834)	Barely mature	33	199	No	1.17	240	100	0.19	0.28
WH6 (2749)	Barely mature	36	191	Yes	1.22	120	> 280	< 0.05	< 0.07
WH3 (2743)	Mature	33	186	Yes	1.32	180	> 220	< 0.06	< 0.08

*Last septum is closely spaced.

†Last septum is septum 1.

‡Based on age difference between septa 1 and 2.

§Angle between septa 1 and 2/rate of septal formation.

||Distance along venter between septa 1 and 2/rate of septal formation.

Table 2 Radiochemical data for septa of specimens of *Nautilus belauensis*

Septum*	$^{210}\text{Po}^\dagger$	$^{210}\text{Po}^\ddagger$	$^{210}\text{Pb}^\ddagger$	$^{210}\text{Po}/^{210}\text{Pb}^\ddagger$	Septum average age (days)
Specimen WH-1					
1	0.56 ± 0.03	0.35 ± 0.04	2.67 ± 0.12	0.13 ± 0.02	28
2	2.32 ± 0.10	2.25 ± 0.11	3.07 ± 0.12	0.73 ± 0.05	260
3	1.96 ± 0.06	1.89 ± 0.07	2.65 ± 0.15	0.71 ± 0.05	240
4	2.60 ± 0.12	2.60 ± 0.12	2.60 ± 0.11	1.00 ± 0.06	> 400
5	2.03 ± 0.10	2.03 ± 0.10	2.21 ± 0.12	0.92 ± 0.07	> 400
6	1.47 ± 0.08	1.47 ± 0.08	1.64 ± 0.08	0.90 ± 0.07	> 400
Specimen WH-2					
1	0.55 ± 0.03	0.45 ± 0.04	1.99 ± 0.18	0.23 ± 0.03	52
2	1.88 ± 0.09	1.82 ± 0.10	2.71 ± 0.10	0.67 ± 0.04	220
3	2.26 ± 0.10	2.16 ± 0.11	3.70 ± 0.28	0.58 ± 0.05	170
4	1.74 ± 0.07	1.74 ± 0.07	1.72 ± 0.08	1.01 ± 0.06	> 400
5	1.88 ± 0.07	1.88 ± 0.07	1.89 ± 0.10	0.99 ± 0.06	> 400
6	2.00 ± 0.15	2.00 ± 0.15	1.78 ± 0.08	1.12 ± 0.10	> 400
Specimen WH-3					
1	1.21 ± 0.05	1.14 ± 0.06	1.91 ± 0.17	0.60 ± 0.06	180
2	2.72 ± 0.12	2.72 ± 0.12	2.93 ± 0.14	0.93 ± 0.06	> 400
3	2.06 ± 0.09	2.06 ± 0.09	1.99 ± 0.14	1.04 ± 0.09	> 400
4	1.92 ± 0.09	1.92 ± 0.09	2.06 ± 0.21	0.93 ± 0.10	> 400
Specimen WH-4					
1	1.71 ± 0.09	1.66 ± 0.10	2.34 ± 0.12	0.71 ± 0.06	240
2	1.39 ± 0.07	1.37 ± 0.08	1.64 ± 0.08	0.83 ± 0.06	340
3	1.85 ± 0.08	1.85 ± 0.08	1.83 ± 0.08	1.01 ± 0.06	> 400
4	1.12 ± 0.06	1.12 ± 0.06	1.22 ± 0.07	0.92 ± 0.07	> 400
Specimen WH-5					
1	0.82 ± 0.04	0.70 ± 0.05	1.68 ± 0.09	0.42 ± 0.04	110
2	1.70 ± 0.06	1.70 ± 0.06	1.75 ± 0.09	0.97 ± 0.06	> 400
3	1.64 ± 0.07	1.64 ± 0.07	1.63 ± 0.12	1.01 ± 0.09	> 400
4	1.68 ± 0.09	1.68 ± 0.09	1.86 ± 0.07	0.90 ± 0.06	> 400
Specimen WH-6					
1	0.96 ± 0.08	0.84 ± 0.09	1.88 ± 0.11	0.45 ± 0.05	120
2	1.86 ± 0.06	1.85 ± 0.07	1.94 ± 0.12	0.95 ± 0.07	> 400
3	1.86 ± 0.08	1.86 ± 0.08	1.79 ± 0.10	1.04 ± 0.07	> 400
4	1.59 ± 0.05	1.59 ± 0.06	1.65 ± 0.08	0.99 ± 0.06	> 400
Specimen WH-7					
1	0.46 ± 0.06	0.25 ± 0.06	4.57 ± 0.22	0.05 ± 0.01	10
2	1.06 ± 0.06	0.98 ± 0.07	2.08 ± 0.09	0.47 ± 0.04	130
3	1.76 ± 0.07	1.72 ± 0.08	2.21 ± 0.12	0.78 ± 0.05	300
4	2.39 ± 0.11	2.35 ± 0.12	2.85 ± 0.15	0.83 ± 0.06	340
5	1.95 ± 0.08	1.95 ± 0.08	1.98 ± 0.09	0.98 ± 0.06	> 400

* Most recent septum = 1.

† Activity at time of initial Po plating.

‡ Value at time of collection of shell. ^{210}Po and ^{210}Pb results are expressed in d.p.m. per g.

septum 32 and a rate of growth of at least 220 days per septum as maturity is approached at septum 36, such an individual would be at least 10 yr old at maturity, an estimate compatible with Saunders' value of ~15 yr¹.

To permit further comparison with mark-recapture observations and growth rate data from aquarium studies, estimates of the angular growth rate (deg per day) and the growth rate along the venter (mm per day) were calculated (Table 1). The angular growth rate was calculated by dividing the angle between the last two septa by the rate of septal formation, and the growth rate along the venter, by dividing the distance along the venter

between the last two septa by the rate of septal formation. In immature animals the angular growth rate ranged from ~0.10 to <0.16 deg per day and the growth rate along the venter from ~0.10 to <0.14 mm per day. By comparison, in specimens 3, 5 and 6, near or at maturity, the angular growth rate ranged from <0.05 to <0.06 deg per day and the growth rate along the venter from <0.07 to <0.10 mm per day. These ventral growth rates are generally lower than those reported in the literature³⁻⁵ but agree well with Saunders' value of <0.12 mm per day¹.

Indeed, the general agreement between the radiometric

growth rates and those obtained from mark-recapture studies supports the two methodologies and permits the conclusion that *N. belauensis* is growing slowly and decreases its rate of growth as it approaches maturity. The differences observed in growth rate between *N. belauensis* and other species of *Nautilus* maintained in aquaria may be artefacts of aquarium conditions, species or habitat specificity, or a reflection of ontogenetic variation.

We thank W. B. Saunders for supplying the specimens of *N. belauensis* and for sharing his observations. This research was supported in part by National Geographic Society grant 2526-82. This is contribution no. 5589 from Woods Hole Oceanographic Institution.

Received 14 October 1983; accepted 31 January 1984.

1. Saunders, W. B. *Paleobiology* 9, 280-288 (1983).
2. Cochran, J. K., Rye, D. M. & Landman, N. H. *Paleobiology* 7, 469-480 (1981).
3. Martin, A. W., Catala-Stuki, I. & Ward, P. D. *Neues Jb. Geol. Paläont. Abh.* 156, 207-225 (1978).
4. Ward, P., Greenwald, L. & Magnier, J. *Paleobiology* 7, 481-493 (1981).
5. Ward, P. D. & Chamberlain, J. *Nature* 304, 57-59 (1983).

Sperm sharing in *Biomphalaria* snails: a new behavioural strategy in simultaneous hermaphroditism

Warton Monteiro, José Maria G. Almeida Jr & Braulio S. Dias

Instituto de Ciências Biológicas, CP: 153081, Universidade de Brasília, 70919 Brasília, DF, Brazil

Simultaneous hermaphroditism includes three reproductive options: self-fertilization, female-acting cross-fertilization, and male-acting cross-fertilization. Sperm sharing is introduced here as a fourth, new option: a cross-fertilized individual shares with other individuals sperm received from its previous crossmate. To share sperm, the individual acts mechanically as a male, but not genetically since the sperm it shares was not produced by it. We show here that sperm sharing occurs in snails of the genus *Biomphalaria* in laboratory conditions.

Biomphalaria are freshwater pulmonate snails widely distributed in Brazil. They are simultaneous hermaphrodites, which reproduce by self-fertilization when in isolation, but whenever two or more individuals are in contact, they copulate and cross-fertilization prevails¹. These snails include species that are vectors of the blood fluke *Schistosoma mansoni*. Our experiments involved samples from six Brazilian populations of three species: *Biomphalaria glabrata*, from Santa Luzia, Minas Gerais State (SL); *Biomphalaria occidentalis*, from Sena Madureira, Acre State (SM); *Biomphalaria tenagophila*, from Núcleo Bandeirante, Distrito Federal (NB); from Joinville, Santa Catarina State (JO); from Jacarepaguá, Rio de Janeiro (JP); and from Belo Horizonte, Minas Gerais State (BH). We had SL, JO and BH albino samples, with total absence of black pigment^{2,3}, and SL, JO, SM, NB, JP wild-type samples. Our laboratory cultures were kept at room temperature (~25°C), fed with romaine lettuce and wheat germ. Our *B. glabrata* laboratory cultures are more than 10 yr old, *B. tenagophila* more than 5 (except NB which is less than 2) and *B. occidentalis*, which is more than 3 yr old. Stocks of albino and wild-type snails were maintained in separate aquaria. Snails with shell diameter ~3 mm were isolated in small aquaria and kept there until the egg-laying stage. This averaged 50 days; the shortest period was 25 days. Afterwards, these snails were observed for at least 1 month, and those that laid eggs regularly were selected for cross-breeding.

Any albino snail cross-fertilized with wild-type sperm lays eggs whose cross-fertilized embryos may be readily identified, before hatching, by their black-spotted eyes. Self-fertilized albino embryos have no black pigment. Eggs were collected

from Styrofoam blocks floating on each aquarium, and transferred to wet cotton in sterilized small vials. About 5 days later, albino and wild-type embryos could be identified. Each experiment consisted of two consecutive matings of snails. Initially, a wild-type snail was paired with an albino, and, afterwards, the latter was paired with another albino snail. This sequence of crosses allowed the transference of sperm from a wild-type snail (donor) to a second albino (recipient), through the first albino (intermediate) (Fig. 1). The various snails were paired in different combinations, resulting in 12 types of trios which are indicated according to the sequence donor-intermediate-recipient, followed by the number of experiments performed with each trio: NB-JO-JO (14), NB-JO-BH (9), NB-BH-BH (13), NB-BH-JO (10), JO-JO-JO (7), JP-JO-JO (7), JP-BH-JO (7), JP-BH-BH (3), SM-BH-BH (7), SM-JO-BH (1), SM-JO-JO (1), SL-SL-SL (18).

All intermediate snails were paired with the recipient only after it was confirmed that they had been inseminated by the donor. All donor × intermediate and intermediate × recipient couples were paired during the day and kept apart overnight. The couples were paired for a period of 2-10 days. However, the intermediate snails were kept isolated after the last donor × intermediate pairing, and before the first intermediate × recipient pairing, for a period of 1-22 days, 10 days on average. Eggs were collected from all albino snails for more than 200 days after the last mating with the wild-type sperm donor. Eggs laid by the intermediate were not recorded during and immediately after intermediate × donor pairing and also not until 10 days after intermediate × recipient pairing. Thus, Table 1 refers to egg laying from days 11-90 after the last pairing with the wild-type sperm donor.

B. tenagophila from Joinville were individually paired, albino × wild type, for 12 h and the eggs of the 16 inseminated albinos were collected daily over 160 days. The frequencies of albino and wild-type eggs were recorded as a control (Table 1, Fig. 2) to compare with the egg production of the albinos (intermediate and recipient) in the trio experiments. For this comparison, only *B. tenagophila* trios were considered.

Our results demonstrate sperm sharing between albino snails acting as intermediate and recipient, since both laid wild-type eggs for more than 100 days after mating (Fig. 2). In 27 out of 97 experiments, transferences of sperm from donor to recipient occurred through an intermediate snail, in the following trios and respective frequencies: NB-JO-JO (6), NB-JO-BH (5), NB-BH-BH (2), NB-BH-JO (3), JP-JO-JO (2), JP-BH-JO (2), SM-JO-BH (1), SM-BH-BH (2), SM-JO-JO (1), SL-SL-SL (3).

Considering *B. tenagophila*-only trios and control crosses, it is possible that control snails have laid, on the average, more eggs than intermediate, or recipient, from the 11th to the 90th day after the last pairing (Table 1, lines 1, 2). But it is evident that wild-type egg production was more abundant among control snails than among intermediate, or recipient, or non-control as a whole (Table 1, lines 3-5). It is also evident that the mean albino egg production by non-control snails was much higher than that of control snails (Table 1, lines 6-8). Albino cross-fertilized eggs can result from intermediate × recipient crosses suggesting that exogenous albino-type spermatozoa may be involved in preferential cross-fertilization, raising the mean albino egg production of non-control snails in relation to controls' (Fig. 2). Although the total egg production of intermediate and recipient snails was, on the average, approximately the same, the production of albino or wild-type eggs, taken separately, was not (Table 1, lines 9-11). This difference may be due to unequal division of the exogenous sperm shared by intermediate and recipient snails. It is possible that sharing experiments may include some spermatozoa loss. In fact, the average wild-type egg production of intermediate plus recipient snails was much smaller than that of controls (Table 1, lines 5 and 11).

In general, the wild-type egg production ceased before that of albino eggs, due to the depletion of the wild-type spermatozoa stock (Fig. 2). Though albino eggs were laid during the whole

Table 1 Test for difference of means (*t*)

	Egg type	Snail	N	$\bar{X}$	s.d.	F	P	<i>t</i>	P
1.	A and Wt	C	11	497.73	241.42	2.65	0.069	1.82	0.091
	A and Wt	I	19	351.68	148.25				
2.	A and Wt	C	11	497.73	241.42	1.63	0.433	1.48	0.156
	A and Wt	R	12	363.42	188.92				
3.	Wt	C	11	441.09	230.19	5.60	0.002	3.75	0.003
	Wt	I	19	167.68	97.24				
4.	Wt	C	11	441.09	230.19	3.84	0.037	4.80	0.000
	Wt	R	12	70.58	117.54				
5.	Wt	C	11	441.09	230.19	4.06	0.003	4.30	0.001
	Wt	I and R	31	130.10	114.24				
6.	A	C	11	56.64	33.41	10.96	0.000	-4.67	0.000
	A	I	19	184.00	110.59				
7.	A	C	11	56.64	33.41	29.89	0.000	-4.40	0.001
	A	R	12	292.83	182.68				
8.	A	C	11	56.64	33.41	20.14	0.000	-5.90	0.000
	A	I and R	31	226.13	149.93				
9.	A and Wt	I	19	351.68	148.25	1.62	0.349	-0.18	0.857
	A and Wt	R	12	363.42	188.92				
10.	A	I	19	184.00	110.59	2.73	0.057	-1.86	0.081
	A	R	12	292.83	182.68				
11.	Wt	I	19	167.68	97.24	1.46	0.459	2.39	0.027
	Wt	R	12	70.58	117.54				

The means refer to albino (A) and/or wild-type (Wt) eggs produced by the control crosses' albino snails (C), intermediate (I), and/or recipient (R) snails of the *B. tenagophila* successful trios. Egg production data were grouped in 80-day periods, from the 11th to the 90th day after the last pairing with the wild-type sperm donor.

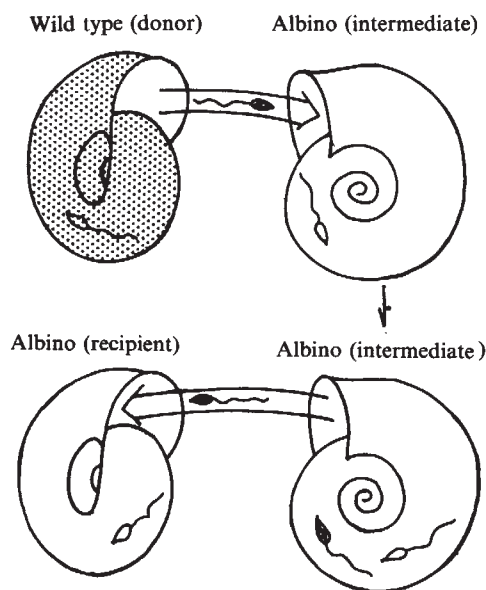


Fig. 1 Sperm transfer from a wild-type (donor) snail to an albino (recipient) through another albino (intermediate) snail.

experiment, a decline in general egg production after day 120 suggests a physiological pause before a new cross to resume high egg production by cross-fertilization. This was not tested.

In 9 out of the 97 trials for sperm sharing, the donor was *B. occidentalis*, and the intermediate and the recipient snails were *B. tenagophila*. Sperm sharing occurred in four of these trials. Crossbreeding between *B. tenagophila* and *B. occidentalis* is common in the laboratory (M. L. F. Dias and W.M., unpublished), though *Biomphalaria* species are generally very well isolated genetically⁴.

Albinism is a mendelian trait useful for tracking wild-type sperm through albino snails. Our method allowed a genetic demonstration of sperm transference. The intermediate snail's reproductive options are: to reproduce by self-fertilization; to receive sperm from the donor, and possibly from the recipient; to transfer its own sperm to the recipient, and possibly to the donor; and to transfer the donor's sperm to the recipient.

The tendency towards cross-fertilization¹ suggests that the snail should have a site or a mechanism that prevents endogenous spermatozoa from mixing with exogenous ones. In fact, in nature *B. obstructa* can harbour sperm from more than one crossmate in a given time⁵. Possibly, all exogenous sperm are stored as a single stock. It seems unlikely that sperm is stored in the spermatheca in *Biomphalaria*⁶; thus, the lumen of the ootestis probably lodges exogenous sperm and guarantees preferential cross-fertilization⁷. However, some self-fertilization does occur during the whole period of cross-fertilized egg production (Fig. 2), perhaps due to some endogenous spermatozoa carried to the ootestis lumen during the exogenous sperm flow through the seminal vesicle.

Sperm sharing seems to be linked to inbreeding and kinship. In fact, natural populations of *Biomphalaria* have some degree of inbreeding, for they commonly live in small water bodies, with little gene flow between populations⁸. It is difficult to dissociate inbreeding from kinship in populations of individuals that reproduce by simultaneous hermaphroditism. As inbreeding and kinship increase in a population, genetic differences between the endogenous and exogenous spermatozoa decrease. Eventually, high levels of inbreeding would virtually eliminate most genetic differences between exogenous and endogenous spermatozoa, and selection pressure to differentiate between endogenous and exogenous spermatozoa would decrease, promoting sperm-sharing behaviour through kin selection^{9,10}. Alternatively, according to Hamilton^{9,11}, inbreeding and low gene flow in natural populations of *Biomphalaria* under local mate competition, favour individuals with a female-biased sex

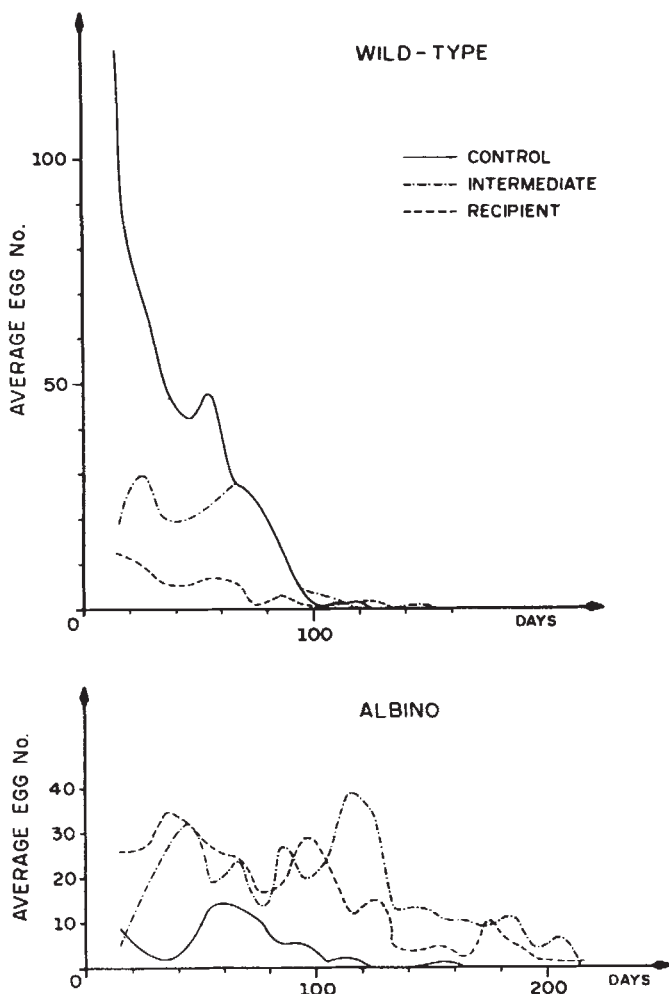


Fig. 2 Distribution of the average frequencies of eggs laid by the albino control snails ($N = 16$), by the intermediate ($N = 21$), and recipient ($N = 21$) of trios with only *B. tenagophila* snails, divided into 10-day class intervals. The data refer to snails with different lifetimes, and do not include eggs laid during the first 10 days after the last pairing with the wild-type sperm donor.

allocation strategy, since it minimizes the competition among sperm of related individuals and maximizes the production of offspring via an increase in egg-cell production; this will favour a female-biased sex ratio. In simultaneous hermaphrodites sex allocation could be biased to female function. Sperm sharing could be a strategy enabling the snail to receive much sperm from its partners, when it is acting as a functional female, while transferring relatively few spermatozoa of its own, when it is acting as a functional male. In this sperm 'commerce', a snail could exchange some of its sperm together with part of the sperm received in a previous cross for a full sperm load from the recipient. This exchange process could be maintained in a population by an evolutionarily stable strategy¹².

We thank Dr Roberto Brandão Cavalcanti for valuable discussions.

Received 5 September 1983; accepted 19 January 1984.

1. Paraense, W. L. *Mems. Inst. Oswaldo Cruz* **53**, 277-291 (1955).
2. Newton, W. L. *Proc. helminth. Soc. Wash.* **21**, 72-74 (1954).
3. Richards, C. S. *Malacol. Rev.* **6**, 49-51 (1973).
4. Paraense, W. L. *Mems. Inst. Oswaldo Cruz* **53**, 325-327 (1955).
5. Mulvey, M. & Vrijenhoek, R. C. *J. Hered.* **72**, 308-312 (1981).
6. Kitajima, E. W. & Paraense, W. L. *J. Morph.* **176**, 211-220 (1983).
7. Paraense, W. L. *Rev. brasil. Biol.* **36**, 535-539 (1976).
8. Narang, N., Narang, S., Monteiro, W. & Almeida, J. M. G. *Jr Rev. brasil. Genet.* **3**, 269-279 (1981).
9. Hamilton, W. D. A. *Rev. ecol. Syst.* **3**, 193-232 (1972).
10. Eberhard, J. J. W. *Q. Rev. Biol.* **50**, 1-33 (1975).
11. Hamilton, W. D. *Science* **156**, 477-488 (1967).
12. Maynard Smith, J. *Am. Scient.* **64**, 41-45 (1976).

Parallel geographical patterns of allozyme variation in two sibling *Drosophila* species

P. R. Anderson & J. G. Oakeshott

Department of Population Biology, Research School of Biological Sciences, Australian National University, PO Box 475, Canberra City, ACT 2601, Australia

The importance of natural selection in determining spatial patterns of allele frequencies remains unclear despite many geographical surveys of allele frequency variation in a variety of organisms¹. Most patterns can be interpreted as a result of either selective or stochastic processes, depending on the assumptions made about population size and migration rate². One line of evidence that could refute the stochastic explanation would be the discovery of parallel patterns of geographical variation for shared polymorphisms in different sibling species³. The present study demonstrates such patterns for the esterase-6 (*Est6*) and phosphoglucosmutase (*Pgm*) polymorphisms of *Drosophila simulans* and *Drosophila melanogaster*. These two species share the same latitudinally correlated patterns of *Est6* allele frequencies in Australasia, North America and Europe, but are more geographically uniform for *Pgm* allele frequencies on all three continents.

Geographical variation in the frequencies of four cosmopolitan chromosome inversions and 10 enzyme polymorphisms have been analysed in populations of *D. melanogaster* covering up to 50° of latitude in Australasia, North America and Europe/Asia⁴⁻¹⁰. Three types of geographical pattern are found. First, all four inversions and six of the enzyme polymorphisms (*Est6*, *Adh*, *Gpdh*, *G6pd*, *Odh* and *Pgd*) show consistent and significant relationships with distance from the Equator in all three zoogeographical zones. Second, two enzyme polymorphisms (*Pgm* and *Idh-NADP*) show no relationship with latitude in any of the three zones. Third, the other two polymorphic loci (*Acph* and *Tpi*) each show a significant latitudinal relationship in Australasia, but not in the other two regions. Oakeshott *et al.*⁹ suggested that random events alone cannot explain the first two patterns, although the interpretation of the third pattern is more difficult.

We have studied the distribution of allele frequencies for analogous polymorphisms in *D. simulans*, a sibling species of *D. melanogaster*. The two species have almost the same cosmopolitan distribution and are often collected together in the field¹¹. Tests under several electrophoretic conditions have indicated that *D. simulans* shares the same common polymorphic electrophoretic variants as *D. melanogaster* for 3 of the 10 enzyme loci analysed in *D. melanogaster*¹²⁻¹⁷: *Est6*, *Pgm* and *Acph*. *D. simulans* is monomorphic for the other seven enzyme systems¹⁴⁻²¹ and has no polymorphic chromosome inversions²². Because of the ambiguous patterns of *Acph* variation in *D. melanogaster*, we have analysed in detail only the patterns for *Est6* and *Pgm* in *D. simulans*. If the world-wide consistent latitudinal relationships of *Est6* allele frequencies and the relative geographical homogeneity of *Pgm* allele frequencies in *D. melanogaster* are simply the result of random events, it is very improbable that they will be repeated in *D. simulans* on all three continents.

An analysis of 80 *D. simulans* population samples for *Est6* and 72 for *Pgm* shows that three electrophoretic variants at each locus are found at overall mean frequencies above 1% (Table 1). Each of the *Pgm* variants has a remarkably similar frequency to its value in *D. melanogaster*, *Pgm-100* clearly being the most common allele in both species. Both *Est6-100* and *Est6-110* are common in the two species, although their relative frequencies differ; *Est6-100* is more common in *D. melanogaster* and *Est6-110* more common in *D. simulans*. The frequency of *Est6-125* is consistently low in both species.

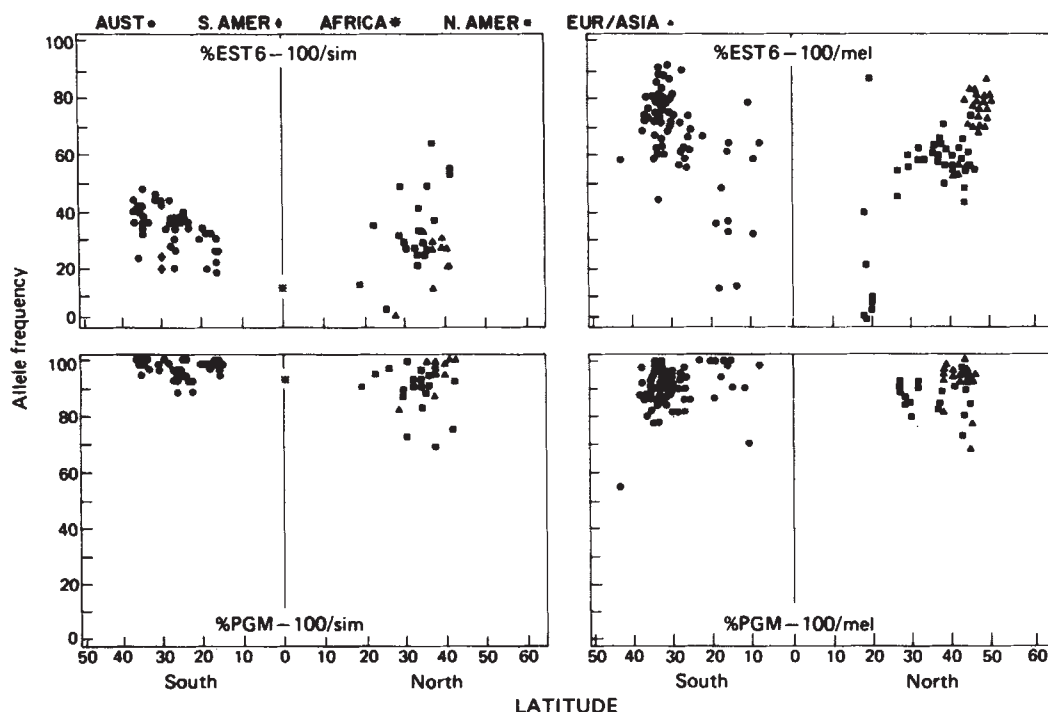


Fig. 1 *Est6-100* and *Pgm-100* frequencies with latitude for *Drosophila simulans* (*sim*) and *D. melanogaster* (*mel*). Data for *D. melanogaster* are from Oakeshott *et al.*⁵.

Only the analyses of variation for the *Est6-100* and *Pgm-100* alleles are presented here. The analysis of the other common allele, *Est6-110*, shows that it has a latitudinal pattern complementary to that of *Est6-100*. The frequencies of the three rare alleles (*Est6-125*, *Pgm-120* and *Pgm-70*) are too low for meaningful analysis.

In Fig. 1 the patterns of latitudinal variation of *Est6-100* and *Pgm-100* frequencies in *D. simulans* and *D. melanogaster* are compared graphically. In each species and in all regions there is a clear relationship between *Est6-100* frequency and latitude. The patterns in the Northern and Southern Hemispheres are mirror-images of each other, and in all three regions the direction of allele frequency change with increasing distance from the Equator is the same in the two species. On average, *Est6-100* frequency increases by approximately 14% for every 10° increase in latitude from the Equator. *Pgm-100* frequency is uniformly high (>85% in most cases, and never <50%) and is not obviously associated with latitude in any region in either species.

Table 1 Mean frequencies of *Est6* and *Pgm* variants

Allele	Mean frequencies* (% $\pm$ s.e.m.)	
	<i>D. simulans</i>	<i>D. melanogaster</i>
<i>Est6</i>		
125	1 $\pm$ 1	2 $\pm$ 2
110	67 $\pm$ 2	36 $\pm$ 2
100	32 $\pm$ 2	62 $\pm$ 2
<i>Pgm</i>		
120	1 $\pm$ 1	3 $\pm$ 1
100	93 $\pm$ 1	89 $\pm$ 1
70	6 $\pm$ 1	8 $\pm$ 1

* For *D. simulans* we obtained 66 of the *Est6* scores and 61 of the *Pgm* scores using electrophoretic methods described elsewhere⁵. The remaining scores were taken from seven other previous reports^{13,15-19,23}. The number of collections analysed and the mean and s.e.m. of the number of genes scored per collection were 47 and 125 ± 12 in Australasia, 20 and 82 ± 16 in North America and 10 and 317 ± 64 in Europe for *Est6*, and 42 and 139 ± 14 in Australasia, 20 and 66 ± 7 in North America and 9 and 311 ± 62 in Europe for *Pgm*. In addition, two populations from South America (Brazil)²³ were scored for *Est6* and one from Africa (Kenya) was scored for both loci. Data for *D. melanogaster* are from Oakeshott *et al.*⁵.

Table 2 gives the partial correlations and standardized regression coefficients of angular transforms of *Est6-100* and *Pgm-100* frequencies on unsigned values of latitude. Both the correlations and regressions show that *Est6-100* frequency is significantly associated with distance from the Equator in both species and all three zones. On average, latitude accounts for about 25% of the variance among samples in *Est6-100* frequency. The only latitudinal association for *Pgm-100* frequency that is significant as both a correlation and regression is for *D. simulans* in Europe. However, only nine European *D. simulans* samples were scored for *Pgm*, and this association is more apparent than real, as it simply results from a relatively low *Pgm-100* frequency in a single, most southerly population.

We have limited data on two South American²³ and one African population of *D. simulans*, and these are consistent with the patterns apparent in the other regions (Fig. 1).

Clinal patterns of variation in a single region can be produced without natural selection²⁴. However, the recurrence of the

Table 2 Geographical variation in *Est6-100* and *Pgm-100* allele frequencies

	<i>D. simulans</i>		<i>D. melanogaster</i> ‡	
	<i>r</i> *	<i>b</i> †	<i>r</i>	<i>b</i>
<i>Est6-100</i>				
Europe	0.66§	0.68§	0.53¶	0.48§
N. America	0.60	0.60	0.31§	0.48
Northern Hemisphere	0.64¶	0.71¶	0.40¶	0.43
Australasia	0.55¶	0.55¶	0.28§	0.28§
Total	0.51¶	0.50¶	0.53	0.57¶
<i>Pgm-100</i>				
Europe	0.80	0.59§	0.35	-0.19
N. America	-0.18	-0.23	0.27	0.46
Northern Hemisphere	-0.02	0.00	0.18	0.20
Australasia	0.27	0.25	-0.15	-0.25§
Total	0.01	0.02	-0.02	-0.02

* *r*, Partial correlations for angularly transformed *Est6-100* and *Pgm-100* allele frequencies with latitude (controlling for longitude).

† *b*, Standardized coefficients for the transformed allele frequencies.

‡ *D. melanogaster* analyses are taken from ref. 5.

§ $P < 0.05$; ¶ $P < 0.01$; ¶¶ $P < 0.001$.

same large-scale latitudinal association for *Est6* in three zoogeographical zones and two species is difficult to explain as solely the result of random processes. The consistent homogeneity of *Pgm* allele frequencies is also not easily explained by non-selective processes; while migration can lead to the convergence of allele frequencies in different populations within a species, this is obviously not possible across species barriers. Thus, although there is a contrast between the geographical patterns of *Est6* and *Pgm* allele frequencies, both patterns probably reflect the action of natural selection.

However, the modes of selection affecting the two loci are likely to be different, and differently influenced by the environment. This may be because *Est6* has externally derived substrates²⁵ which may vary qualitatively or quantitatively with latitude, and *Pgm* is involved in central metabolism²⁶ with substrates more likely to be consistent across latitudes.

For both *Est6* and *Pgm* it is likely that some amino acid replacements are not detected by electrophoresis^{23,27,28}. However, such cryptic variation would tend to obscure (rather than produce) any recurrence of systematic regional patterns across species.

It is possible that the similar geographic patterns of the shared polymorphisms in the two species are the result of selection acting on genes closely linked to the enzyme loci, rather than on those loci themselves. If this is so, the identity of such genes and their disequilibrium with the loci studied would presumably have to be similar in each species. Similar disequilibrium could be established in both species if it ante-dated speciation (at least 10^6 generations ago²⁹). However, it seems likely that most of this disequilibrium would have subsequently decayed, unless maintained by epistatic selection. The loci studied could participate directly in any such epistatic selection or be embedded among other epistatically-selected genes³⁰. Except for the latter case then, we consider that the geographical patterns described here are most likely to be due to selection involving *Est-6* and *Pgm*.

We thank I. A. Boussy, S. Easteal, J. B. Gibson and W. R. Knibb for valuable discussions, D. A. Willcocks for technical assistance and F. M. Cohan, R. Anderson and W. R. Knibb for assisting in the collection of *Drosophila*.

Received 25 October 1983; accepted 26 January 1984.

1. Nevo, E. *Theor. Pop. Biol.* **13**, 121–177 (1978).
2. Lewontin, R. C. *The Genetic Basis of Evolutionary Change* (Columbia University Press, New York, 1974).
3. Harrison, R. G. *Nature* **266**, 168–170 (1976).
4. Anderson, P. R. in *Genetic Studies of Drosophila Populations* (eds Gibson, J. B. & Oakeshott, J. G.) 237–250 (Australian National University Press, Canberra, 1981).
5. Oakeshott, J. G., Chambers, G. K., Gibson, J. B. & Willcocks, D. A. *Heredity* **47**, 385–396 (1981).
6. Oakeshott, J. G. *et al. Evolution* **36**, 88–96 (1982).
7. Oakeshott, J. G., Chambers, G. K., Gibson, J. B., Eanes, W. F. & Willcocks, D. A. *Heredity* **50**, 67–72 (1983).
8. Oakeshott, J. G., Gibson, J. B. & Willcocks, D. A. *Theor. appl. Genet.* **65**, 191–196 (1983).
9. Oakeshott, J. G., McKechnie, S. W. & Chambers, G. K. *Genetica* (in the press).
10. Knibb, W. R., Oakeshott, J. G. & Gibson, J. B. *Genetics* **98**, 833–847 (1981).
11. David, J. R. & Tsacas, L. *C. r. somm. Séanc. Soc. Biogeogr.* **57**, 11–26 (1980).
12. Eisses, K. T., van Dijk, H. & van Delden, W. *Evolution* **33**, 1063–1068 (1979).
13. Triantaphyllidis, C. *J. Hered.* **64**, 69–72 (1973).
14. O'Brien, S. & MacIntyre, R. *Am. Nat.* **103**, 97–113 (1969).
15. Triantaphyllidis, C. D., Panourgias, J. N., Scouras, Z. G. & Ioannides, G. C. *Genetica* **51**, 227–231 (1980).
16. Triantaphyllidis, C. D., Scouras, Z. G., Panourgias, J. N. & Ioannides, G. C. *Genetica* **58**, 129–136 (1982).
17. Cabrera, V. M., Gonzalez, A. M., Larruga, J. M. & Guillon, A. *Genetica* **59**, 191–201 (1982).
18. Kojima, K.-I., Gillespie, J. & Tobari, Y. N. *Biochem. Genet.* **4**, 627–637 (1970).
19. Steiner, W. W. M., Sung, K. C. & Paik, Y. K. *Biochem. Genet.* **14**, 495–506 (1976).
20. Berger, E. M. *Genetics* **66**, 677–683 (1970).
21. Ohnishi, S., Leigh Brown, A. J., Voelker, R. A. & Langley, C. H. *Genetics* **100**, 127–136 (1982).
22. Ashburner, M. & Lemeunier, L. *Proc. R. Soc. B193*, 137–157 (1976).
23. de Albuquerque, C. M. R. & Napp, M. *Genetics* **98**, 399–407 (1981).
24. Endler, J. A. *Geographic Variation, Speciation and Clines* (Princeton University Press, New Jersey, 1977).
25. Zera, A. J., Koehn, R. K. & Hall, J. G. in *Comprehensive Insect Physiology, Biochemistry and Pharmacology* Vol. 10 (eds Kerkut, G. A. & Gilbert, L. I.) Ch. 18 (Pergamon, New York, 1983).
26. Dickinson, W. J. & Sullivan, D. T. *Gene-Enzyme Systems in Drosophila* 130–136 (Springer, Berlin, 1975).
27. Cochran, B. J. & Richmond, R. C. *Genetics* **93**, 461–478 (1979).
28. Trippa, G., Catamo, A., Lombardozzi, A. & Cicchetti, R. *Biochem. Genet.* **16**, 299–305 (1978).
29. Leigh Brown, A. J. & Ish-Horowitz, D. *Nature* **290**, 677–682 (1981).
30. Franklin, I. R. & Lewontin, R. C. *Genetics* **65**, 701–734 (1970).

Disruption of cortical activity prevents ocular dominance changes in monocularly deprived kittens

C. Shaw & M. Cynader

Departments of Psychology and Physiology and Biophysics, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1

Abundant evidence now indicates that atypical visual exposure early in the life of cats and primates can cause profound alterations in cortical organization¹. In particular, it has been shown that preventing the use of one eye for vision early in life results in a marked shift of ocular preference among neurones of kitten visual cortex in favour of the exposed eye^{1–3}. The cellular mechanisms underlying these alterations remain uncertain, but much recent attention has focused on the possible role of pharmacological agents in modifying cortical plasticity, with particular reference to catecholamines. These experiments, which have shown that agents which modify cortical nonadrenaline levels can alter the degree of cortical plasticity^{4–7}, do not specify the mechanism of action, and leave open the possibility that other neurotransmitter systems may also be involved in cortical modifiability. We now report that chronic intracortical administration of L-glutamate during a period of monocular vision imposed on young kittens largely prevents the ocular dominance shift which normally occurs under these circumstances.

Eight kittens were monocularly deprived by eyelid suture for 1–2 weeks, beginning 5–7 weeks postnatally (see Table 1). At the time of eye closure cannulae constructed from 30-gauge syringe needles were stereotactically implanted into the visual cortex of both hemispheres. The tips of cannulae were positioned 0.5–0.85 mm below the cortical surface, with the bevelled opening facing anteriorly from the implant site. The cannulae were connected via PE tubing to osmotic minipumps (Alza Corporation, models 2001 ($1.0 \mu\text{l h}^{-1}$) or 2002 ($0.5 \mu\text{l h}^{-1}$)) containing either L-glutamate (0.4–0.5 M) in saline (pH 7.3–7.7) or saline carrier solution alone⁸. Glutamate is known to increase the firing of cortical neurones, but not afferent fibres in cat visual cortex⁹. In the following discussion, 'control' will denote the saline-infused cortex, while 'experimental' will denote the opposite glutamate-infused cortex. Three of the eight kittens were prepared and recorded using a blind procedure in which the identity of the control and experimental hemispheres was revealed only on completion of all the experiments. All of the above kittens were placed in the dark for 2 or more days at the end of the infusion period to ensure that the glutamate infusion had been exhausted. Four additional kittens were allowed binocular exposure, and evoked and spontaneous activity was examined in neurones of the visual cortex during glutamate infusion. Details of surgical preparation and unit recording techniques have been published previously^{8,10}.

Figure 1a–c illustrates results from a typical kitten which was allowed monocular exposure during glutamate infusion. In the control hemisphere (Fig. 1a), one observes the familiar loss of binocular influence and a shift in the ocular dominance distribution in favour of the exposed eye. On the experimental side (Fig. 1c), binocular influence is less affected and only a small trend in favour of the exposed eye is observed. At distances >3 mm from the implant site in the experimental hemisphere (Fig. 1b) the ocular dominance distribution is shifted towards the open eye, but still not as much as in the control hemisphere. Other investigators have reported a much smaller range (up to 200 μm) over which glutamate usually alters neural activity¹¹. However, these authors used iontophoretic techniques and a much smaller quantity of glutamate over much shorter periods of time.

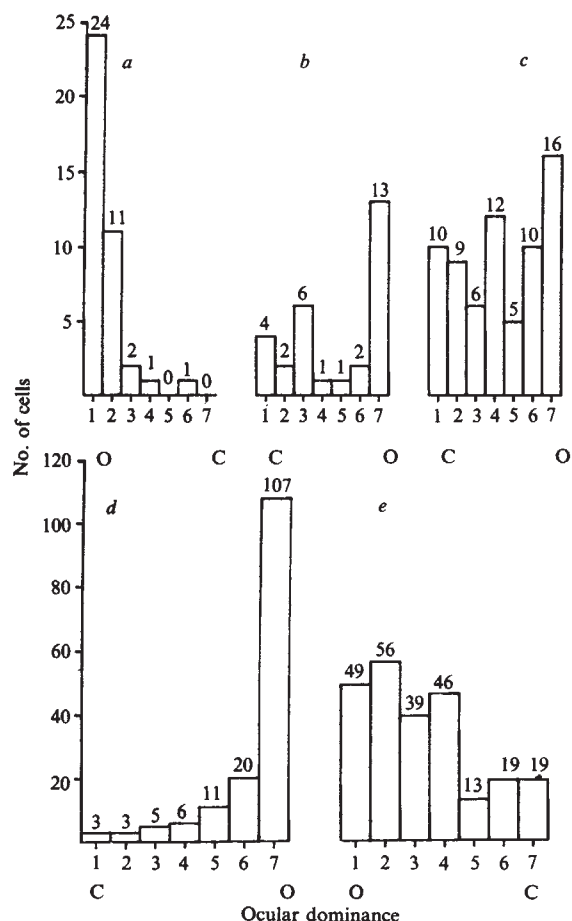


Fig. 1 *a-c*, Distribution of ocular dominance among cortical cells in a representative kitten (GK-10) subjected to monocular eyelid suture at 5 weeks of age. Osmotic minipumps containing L-glutamate or saline were implanted in a blind procedure. *d, e*, Combined results from all eight kittens subjected to monocular eyelid suture at 5-7 weeks postnatally. Conventions described below. *d*, Control hemispheres with saline infusion. *e*, Experimental glutamate-infused hemispheres. The ocular dominance distributions plot the incidence of cells (on the ordinate) against a seven-point scale of ocular dominance devised by Hubel and Wiesel¹⁹. On this scale the numbers from 1 to 7 represent a spectrum of ocular dominance from complete dominance by the contralateral eye (group 1) through equal input from the two eyes (group 4) to complete dominance by the ipsilateral eye (group 7). *a* Shows that in the control hemisphere, which received saline infusion only, most cells are driven well only via input from the contralateral eye which remained open through the deprivation period. In all kittens the difference in ocular dominance distributions between the control and experimental cortices became extremely clear as the experiment proceeded. In the animals prepared and recorded as double-blind controls, it was easy to predict, within a very short time, which treatment each hemisphere had received. Based on this prediction, we recorded at varying distances from the implant site in this kitten in order to gain a clearer idea of the range over which glutamate infusion prevented the ocular dominance shift. *c* Shows that the effects of monocular deprivation are considerably less pronounced than in the other visual cortex into which saline was infused. Recordings here were made from 0.25 to 2.5 mm from the infusion site. The centre ocular dominance distribution (*b*) represents recordings made in the experimental hemisphere, but 3-4 mm from the infusion site. The effects of monocular deprivation appear more pronounced here than near the site of infusion. In *e* the graph for the control hemisphere is plotted as though the contralateral eye was always sutured and therefore that the ipsilateral eye was always sutured for recordings in the experimental hemisphere. This procedure was used in five of the eight animals studied. For one of the three animals studied with the blind procedure described in the text, the experimental hemisphere happened to be contralateral to the sutured eye, but we have plotted all the results as though the contralateral eye was always the sutured eye for recording the control hemisphere. Note that the naturally-occurring trend for contralateral eye dominance¹⁹ would minimize the glutamate effects reported here. In all histograms, C is the closed and O the open eye, respectively.

Table 1 Ocular dominance in monocularly deprived kittens

Cat code	Age at eyelid suture (days)	Age at recording (days)	Period of infusion and monocular visual exposure (days)	No. of cells§	% Ocular dominance groups 2-6	% Open eye only	% Open eye dominated groups 1-3
GK-1	49	58	6	Control 22	45	50	73
				Expt 58	88	12	61
GK-2	50	69	13	Control 14	50	50	79
				Expt 30	70	17	50
GK-5	35	55	14	Control 12	0	100	100
				Expt 25	60	28	60
GK-7*†	37	68	14	Control 16	6	94	100
				Expt 0	—	—	—
GK-8†	41	60	14	Control 18	28	72	95
				Expt 0	—	—	—
GK-9*	41	66	14	Control 15	33	63	100
				Expt 37	78	19	65
GK-10*	35	84	14	Control 32	25	75	94*
				Expt 73	60	26	64*
GK-11	35	65	14	Control 26	35	58	81
				Expt 18	72	27	94
Totals:							
Control hemispheres				155	29	69	89
Experimental hemispheres				241	72	20	60
Experimental hemispheres remote from infusion site‡ (>3 mm)				24	42	42	50

* Prepared and recorded in double-blind procedure.

§ Does not include cells recorded >3 mm from the infusion site (see 'Totals') or cells recorded during infusion.

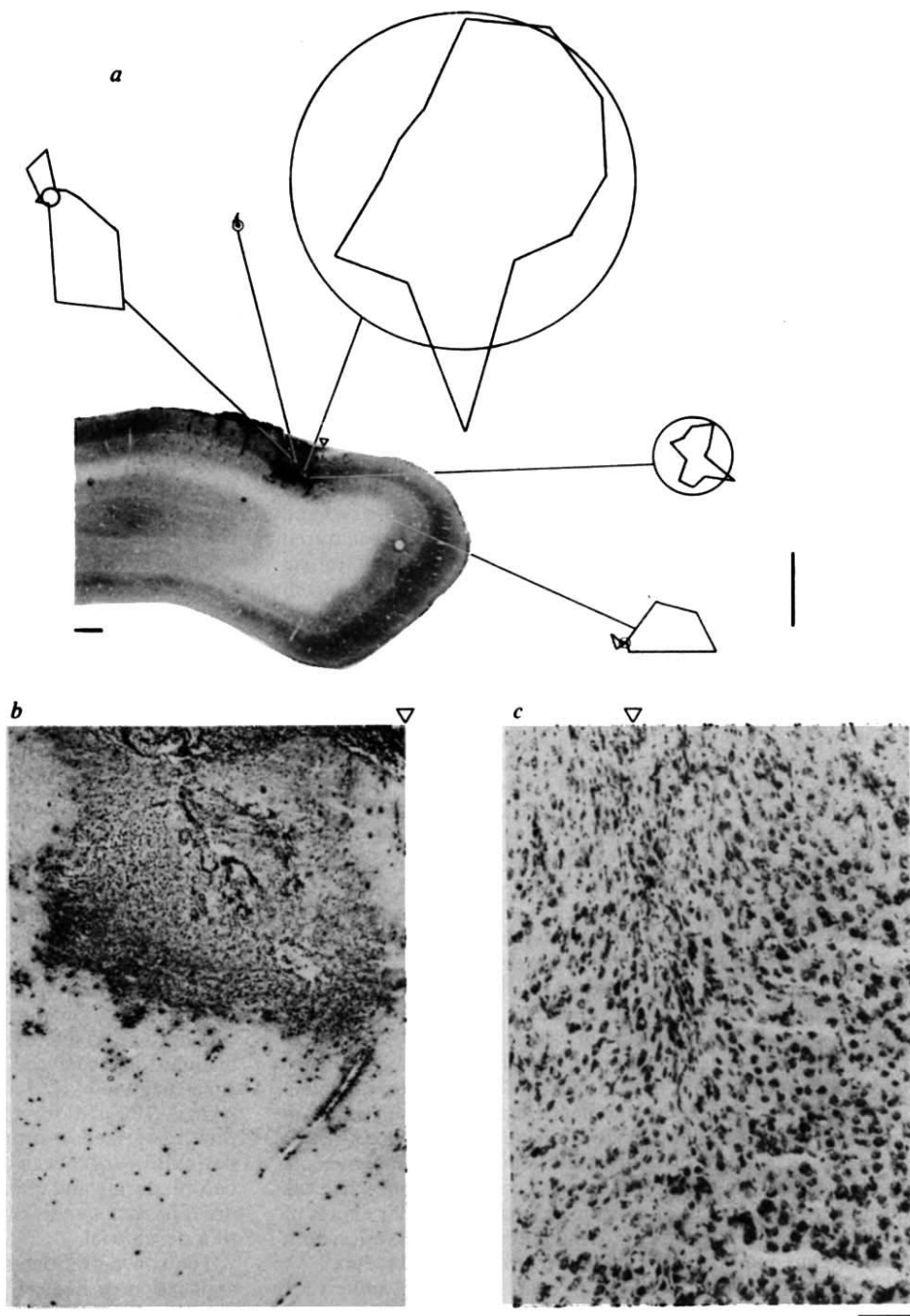
† Sample too small for Mann-Whitney *U*-test.

‡ Animal died before unit recording could be made in the other hemisphere.

|| Significantly different from control *P* < 0.005.

* The glutamate-infused cortex was ipsilateral to the open eye.

Fig. 2 The chronic effects of 12 days of glutamate infusion (0.5 M , $0.5\text{ }\mu\text{l h}^{-1}$) are examined by single unit recording and cytochrome *c* oxidase histochemistry. **a**, A parasagittal section ($20\text{ }\mu\text{m}$) containing the infusion site, and an electrode track passing through this site ($A \rightarrow P$), was processed for cytochrome *c* oxidase activity. The enzyme reaction products are darkest in the area surrounding the infusion site, especially in layers IV and V. Scale bar, 1 mm . Ventral, down; anterior, left. The open triangle indicates the implant site. Serial reconstruction of the region of heightened cytochrome oxidase activity suggests that this patch covered an area of $2.5\text{--}3\text{ mm}$ ($A\text{--}P$) by $1.0\text{--}1.2\text{ mm}$ ($M\text{--}L$). Similar cytochrome *c* oxidase results were obtained in another kitten. These results agree well with the physiologically determined infusion area (Fig. 1*b, c*). In contrast, saline infusion alone did not engender patches of heightened cytochrome *c* oxidase activity. Bars or gratings at six different orientations, spaced 30° apart, were moved under computer control during single unit recording. For each orientation, there were two directions of motion. In this experiment a 2.6° period grating was used, moved at $1\text{--}4\text{ deg s}^{-1}$ (four repetitions). The orientation selectivities of five representative neurones encountered during the penetration are shown in polar plots. In each case, the length of the line in a given direction indicates the strength of response for movement in that direction. For example, if the longest vector points directly downward (270°) it indicates that the preferred direction of movement for a horizontally oriented grating is down. Scale bar, 25 spikes. The open circles show spontaneous activity. **b**, Same section as **a**, but at a higher magnification. Scale bar, $100\text{ }\mu\text{m}$. The intensely stained dark spots in various areas of the section are red blood cells. **c**, an adjacent section from the same animal has been stained with cresyl violet. An examination of neurones in regions corresponding to the heightened cytochrome *c* activity shows no pronounced pathology. Scale bar, $100\text{ }\mu\text{m}$. Ventral, down; anterior, left. The open triangle indicates the implant site.



Essentially similar results were found in all eight kittens, including the subjects studied with the double-blind technique described above. Figure 1*d, e* and Table 1 summarize the results from all kittens. As shown for the individual animal, binocular convergence on single cells was always lower in the control hemisphere and the ocular dominance shift towards the open eye was significantly lower in the experimental than control cortices (Mann-Whitney *U*-test, $P < 0.005$).

Despite the relatively long period of drug administration, the general properties of cortical cells seemed unaffected once infusion was discontinued. The incidence of cells displaying orientation and/or directional selectivity showed little or no difference between control and experimental hemispheres. Velocity preferences, receptive field sizes and types, response strengths and spontaneous activity were all similar in the experimental and control hemispheres. Additionally, there were no marked differences in the numbers of non-excitable cells recor-

ded in either hemisphere. This latter result is perhaps surprising considering the known excitotoxic and pathological effect of glutamate on cortical neurones¹². While some deleterious effects of glutamate might have been expected at the implant site, it was difficult to distinguish these from possible pathology caused by the implant itself (see Fig. 2*c*).

To evaluate the significance of the previous results, we examined the effects of chronic glutamate infusion on neural activity during the actual infusion. Four kittens were prepared essentially as described above, but examined after 4–13 days of glutamate infusion. These kittens had only unilateral implants and were not deprived by eyelid suture. Electrode penetrations were angled to intersect the infusion site. All units encountered (minimum interval between cells $100\text{ }\mu\text{m}$, $n = 75$) were first plotted using a hand-held projector. Orientation selectivity, response amplitude and spontaneous activity were then examined using bars or gratings moved under computer control.

The results were essentially the same in all four animals. On entering the cortex relatively far from the infusion site, most cells encountered had low spontaneous activity and showed strong responses with marked orientation and direction selectivity. Next a zone of very low spontaneous activity was encountered in which the units' evoked activity was also unusually weak. As the penetration continued, a zone of high spontaneous activity was entered. In this area, it was difficult to distinguish spontaneous from evoked activity¹¹. Leaving this area, the effects were reversed. The electrode first passed through a patch of poorly responsive cells before finally encountering cells with normal spontaneous and evoked activity. These patterns of activity were not encountered in regions >3 mm from the infusion site.

Figure 2a and b illustrate these results, showing at different magnifications a parasagittal section of the visual cortex which had been processed to detect the activity of cytochrome c oxidase, a mitochondrial enzyme whose presence reflects activity in the area of cortex being stained¹³. The cytochrome c oxidase reaction in this section shows enhanced activity adjacent to the infusion site, particularly in layers IV and V. The evoked and spontaneous activity of representative neurones encountered during the penetration are illustrated by orientation tuning curves. It is clear from the cytochrome c oxidase stain and from the neural responses that the glutamate infusion markedly disrupts the normal responsivity of cortical neurones. Figure 2c shows an adjacent section stained with cresyl violet. An examination of the region corresponding to the heightened cytochrome c oxidase activity shows normally stained neurones.

The results show that chronic glutamate infusion can prevent shifts in ocular dominance in the region of cortex in which activity is altered by the infusion. These results support the notion that normal postsynaptic cortical activity is an important determinant of cortical plasticity. The effects of glutamate are likely to occur primarily at the level of cortical neurones⁹, leaving thalamic inputs relatively unaltered (but see ref. 14) and so they suggest that the ability of the cortical neurone to respond effectively to its input is important for plasticity.

Recently, there has been much interest in the possibility that plasticity in kitten visual cortex is regulated by various catecholamines, notably noradrenaline⁴⁻⁷. While catecholamines may indeed have a role in cortical plasticity, the results reported here suggest caution in the interpretation of these findings. It may be that the chronic depletion of noradrenaline alters postsynaptic activity and hence signal to noise ratios¹⁵ in a manner similar to that which we have reported here. If this is so, the mechanism by which catecholamine-depleting agents act on cortical plasticity may be very different from what has been suggested. It is interesting in this light that clear effects of iontophoretically applied noradrenaline can be observed on both spontaneous and evoked neural activity in cat visual cortex¹⁶⁻¹⁸.

We thank M. Feran, V. Baker, S. Kirby and C. Lambert for technical assistance. This study was supported by NRSA award 1FYEY05393-01 and Killam Postdoctoral Fellowship (to C.S.) and grants from the MRC (MT-5201) and NSERC (A9939) (M.C.).

Received 26 August 1983; accepted 20 January 1984.

1. Movshon, J. A. & Van Slysters, R. C. A. *Rev. Psychol.* **32**, 477-522 (1981).
2. Hubel, D. H., Wiesel, T. N. & Levay, S. *Phil. Trans. R. Soc. B278*, 131-163 (1977).
3. Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **206**, 419-436 (1970).
4. Kasamatsu, T. & Pettigrew, J. D. *J. comp. Neurol.* **185**, 139-161 (1979).
5. Kasamatsu, T., Pettigrew, J. D. & Ary, M. *J. comp. Neurol.* **185**, 163-181 (1979).
6. Daw, N., Rader, R. K. & Robertson, T. W. *Soc. Neurosci. Abstr.* **7**, 673 (1981).
7. Daniels, J. D., Ellis, M. K., Bianco, S. A., Garrett, M., Nelson, S. B. & Schwartz, M. *Soc. Neurosci. Abstr.* **7**, 673 (1981).
8. Shaw, C. & Cynader, M. *Soc. Neurosci. Abstr.* **7**, 722 (1981).
9. Hess, R. & Murata, K. *Expl. Brain Res.* **21**, 285-297 (1974).
10. Cynader, M., Berman, N. & Hein, A. *Expl. Brain Res.* **25**, 139-156 (1976).
11. Hess, R., Negishi, K. & Creutzfeldt, O. *Expl. Brain Res.* **22**, 415-419 (1975).
12. Van Harreveld, A. & Fikova, E. *Expl. molec. Path.* **15**, 61-81 (1971).
13. Wong-Riley, M. *Brain Res.* **171**, 11-28 (1979).
14. Thieffry, M., Bruner, J. & Personne, P. *J. Physiol., Paris* **75**, 635-639 (1979).
15. Woodward, D. J., Moises, H. C., Waterhouse, B. D., Hoffer, B. J. & Freedman, R. *Fedn. Proc.* **38**, 2109-2116.
16. Reader, T. A. *Experientia* **34**, 1586-1587 (1978).
17. Reader, T. A., Ferron, A., Descarries, L. & Jasper, H. H. *Brain Res.* **160**, 217-229 (1979).
18. Kasamatsu, T. & Heggelund, P. *Expl. Brain Res.* **45**, 317-327 (1982).
19. Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **160**, 106-154 (1962).

Release of endogenous Zn²⁺ from brain tissue during activity

S. Y. Assaf & Shin-Ho Chung

National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

The role of divalent transition metal ions in neural function is poorly understood. In excess, these ions are associated with neurological disorders such as Wilson's disease¹, Pick's disease² and epileptic seizures^{3,4}. We suggest that zinc ions, which are contained in nerve terminals, are extruded into the extracellular space during neuronal activity. Excessive levels of zinc may be released during intense neuronal activation, and contribute to the paroxysm and toxic damage observed. Zinc ions are contained in high concentrations in mossy fibres of the hippocampal formation⁵, and it is the postsynaptic neurones of these fibres which are most susceptible to the toxic effects of kainic acid⁶, a potent convulsant, or to chronic exposure to organometallic compounds⁷. Here we demonstrate for the first time that Zn²⁺ is released into the extracellular space during excitation of hippocampal slices.

For each experiment, eight transverse slices of rat hippocampus, each 150 µm thick, were maintained at 36°C in a specially constructed chamber⁸, which was continuously perfused with normal solution at a rate of 1-3 ml min⁻¹. The composition of the normal perfusing medium was (mM): 136.0 NaCl, 2.56 KCl, 1.18 KH₂PO₄, 1.18 MgSO₄, 1.8 CaCl₂, 10.0 D-glucose and 16.0 NaHCO₃. KCl concentrations in the solution were increased by KCl/NaCl substitution. To minimize the concentration of Zn²⁺ in the solution, we used the 'Aristar' grades of NaCl, D-glucose and KCl, the 'Analar' grades of the remaining chemicals (BDH Chemicals, Ltd), and double-deionized water. The concentration of Zn²⁺ in the samples was determined with an atomic absorption spectrophotometer (Pye Unicam SP9), using the principal absorption line (213.9 nm). The average Zn²⁺ content of the normal medium was 17.4 ± 0.9 parts in 10⁹ (p.p.b.). We determined the protein content of the eight slices used in the experiments to be 4.4 mg, from which we estimated their wet weight to be 45 mg. After 1 h of perfusion, the normal perfusing medium was replaced with medium containing an increased concentration of K⁺ (either 12.5, 23.8 or 50 mM) or kainic acid (KA; 1-100 µM) for a period of 10 or 12 min. The tissues were then re-washed in normal medium, and additional samples of perfusate were collected for another hour. The samples collected before, during and after the application of KA or high K⁺ were dried *in vacuo*, redissolved in 1 ml of 5 M HCl, and transferred to a plastic vial.

The concentration of Zn²⁺ in perfusate collected during the application of high K⁺ was higher than that collected during control periods. Figure 1a shows that the largest extrusion of Zn²⁺ occurred during the first 6 min of stimulation, when electrophysiological recordings indicated that the neurones were displaying abnormal electrical activity characterized by a transient enlargement of evoked responses (Fig. 1a, insets) followed by a period during which responses could not be evoked (not shown), probably due to massive depolarization of the cells. The concentration of Zn²⁺ in the fluid perfusing the tissues increased to a maximum of 30.7 ± 3.2 × 10⁻⁹ during application of 23.8 mM K⁺ and then declined gradually to the background level. The average volume of fluid collected for each 3-min sample was 3.98 ± 0.12 ml. When the tissues were perfused with the normal solution again, the electrical activity slowly returned to normal. There was a small, but statistically significant, decrease of the Zn²⁺ content in the samples collected during the recovery period, suggesting that tissues were accumulating Zn²⁺ contained in the perfusing media. All but two of the 21 experiments using K⁺ concentrations of 23.8 or 50 mM showed a similar increase of the Zn²⁺ concentration in the perfusate collected during the application of a high K⁺ solution. By calcu-

lating the total amount of Zn^{2+} released, and taking the hippocampal Zn^{2+} content of the rat to be 16 p.p.m.^{9,10}, we have estimated that ~18% of Zn^{2+} contained in the hippocampal slices was extruded during the application of high K^+ .

This release of Zn^{2+} by the tissues during hyper-excitation is calcium-dependent. In six experiments, hippocampal tissues were perfused with medium containing a low Ca^{2+} concentration and high Mg^{2+} concentration for 1 h. This treatment, as expected, abolished evoked synaptic potentials (see inset labelled 0 in Fig. 1b). The tissues were then perfused with a solution containing high K^+ for 8 min, after which they were allowed to recover in the normal solution. In all these experiments, a pulse of high potassium failed to release Zn^{2+} into the fluid superfusing the tissues. The measurements obtained from one such experiment are shown in Fig. 1b.

As kainic acid is known to produce prolonged excitation¹¹ and toxic damage of hippocampal neurones⁶, we perfused our slices with this toxin. A large amount of Zn^{2+} was released by hippocampal tissues stimulated by KA. After 1 h of incubation in the normal solution, tissues were similarly perfused with a solution containing KA 1, 10 or 100 μM except that only four samples, each lasting 5 min, were collected and concentrated 10-fold as described above. The average volume of fluid collected during each 5-min period was 13.5 ± 0.7 ml. The results of 14 experiments, each containing eight hippocampal slices, are summarized in Fig. 1c. The mean difference in the concentration of Zn^{2+} between the control value and after the first 5 min of KA application is 5.6 ± 1.5 p.p.b. whereas that between the control value and the second 5 min of KA application is 4.2 ± 1.3 p.p.b. There was a small but significant decrease (2.2 ± 1.1 p.p.b.) in the concentration of Zn^{2+} in the perfusate collected 1 h after application of KA, again suggesting that the slices were taking up some of the Zn^{2+} contained in the normal medium. Analyses of the time course of this release showed that the period of maximal extrusion of Zn^{2+} coincides with the period during which the tissues were hyper-excitable (see inset labelled +5 in Fig. 1c).

The loss of Zn^{2+} from the hippocampus was also demonstrated in intact animals. Nineteen rats were injected intraperitoneally (i.p.) with KA at 19.4 mg per kg to produce convulsions for 2 h as described previously¹², and their hippocampal and serum zinc content were compared with 19 matched control rats which received i.p. injections of saline. The hippocampus of the convulsed rats contained $9.7 \pm 2.7\%$ less Zn^{2+} than that of the normal rats. On the other hand, the serum of the convulsed rats contained $12.6 \pm 3.5\%$ more Zn^{2+} than the control samples. Thus, we conclude that hippocampal Zn^{2+} is depleted during epileptic episodes.

We have considered, and eliminated, several possible sources of artefacts. The observed increase in the concentration of Zn^{2+} during application of KA and high K^+ could not have been due to contamination from the chemicals, as if this were the case, the concentration of Zn^{2+} in all samples collected during the application of the solutions would have been high. This does not accord with our findings. Moreover, efflux of Zn^{2+} was not detected in a Ca^{2+} -free medium or when the doses of the convulsants used were below 12 mM K^+ or 1 μM of KA.

That endogenous Zn^{2+} is released from brain tissue by KA and by increased extracellular potassium is a novel finding which has several implications. The hippocampal formation of rat contains 16.2 ± 0.3 p.p.m. of Zn^{2+} (refs 9, 10). Our calculation reveals that 18% of the total zinc contained in this tissue is extruded into the perfusing medium when the perfusing solution contains 23.8 mM KCl. If we assume that the extracellular space occupies 15% of the brain, the local concentration of Zn^{2+} at the peak of convulsive activity must reach almost 300 μM . Such a transient increase in the extracellular concentration of Zn^{2+} could have a profound effect on neuronal function. From the known complexing behaviour of zinc ions, we can deduce that they will bind with organic ligands and perhaps interfere with transmitter-receptor interaction. However, the postsynaptic actions of Zn^{2+} on hippocampal neurones are unknown and we

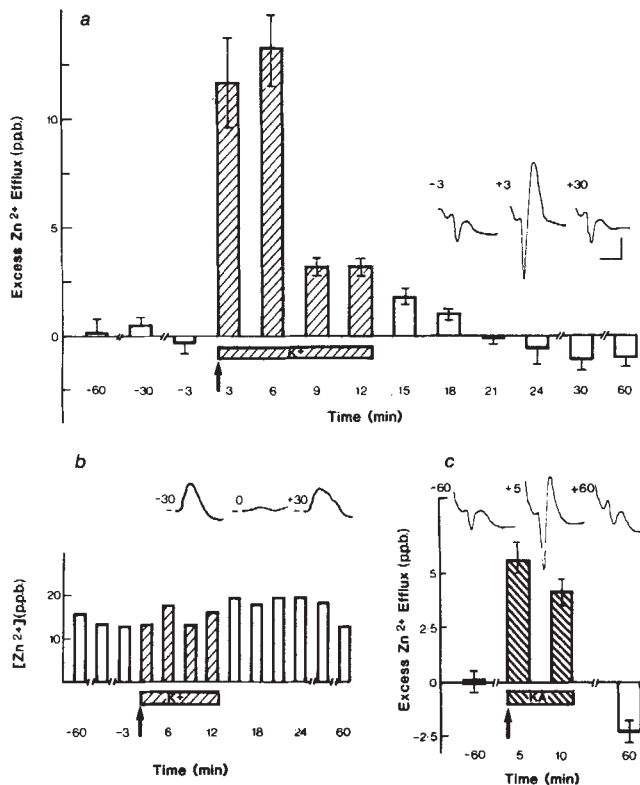


Fig. 1 Release of Zn^{2+} from hippocampal slices in vitro. *a*, The Zn^{2+} content of the perfusate (allowing for the concentration factor) collected before, during and after exposure to a high K^+ concentration (23.8 mM) is compared with that of the normal solution (17.4 ± 0.9 p.p.b.). Samples of perfusate were collected at the time indicated on the abscissa. At the arrow, the perfusing medium was changed from the normal solution to a solution containing high K^+ . The bars represent the average differences in the Zn^{2+} concentration between the control solution and the perfusates ($n = 7$). Error bars represent 1 s.e.m. The three traces in the inset are evoked responses recorded from the CA3 regions following electrical stimulation of the mossy fibre bundle. Note that the evoked response obtained 3 min after the application of high K^+ is higher in amplitude than the two control responses. Calibration bars for all traces: 5 mV and 5 ms. *b*, The bars represent the concentrations of Zn^{2+} in perfusate collected in one typical experiment, in which the perfusates were normal medium (at -60 and +60 min) and low Ca^{2+} /high Mg^{2+} (between -30 and +30 min). The K^+ concentration of the latter perfusate was increased at the arrow and continued for the time indicated by the hatched bars. The synaptic potential in the low Ca^{2+} solution was abolished (see inset) and did not recover until the normal medium was perfused again. *c*, The differences in Zn^{2+} content between the perfusates collected before, during and after the application of 100 μM KA (averages of 14 experiments). Each collection period lasted for 5 min. The insets indicate that the evoked response was enhanced during KA application (+5 min).

are now performing experiments to investigate these actions. Previous work on lobster muscle and mammalian cortical neurones indicates that the presence of Zn^{2+} in low concentrations in the incubating medium interferes with γ -aminobutyric acid-mediated responses^{13,14}. Also, a massive efflux of Zn^{2+} in a confined extracellular space may alter the pH of the microenvironment, leading to cellular damage. The CA3 hippocampal neurones, which are innervated by the mossy fibre bundle containing the highest concentration of Zn^{2+} (ref. 5), are among the first group of cells destroyed by KA⁶. The possibility that the toxic effects of this drug are due to the release of Zn^{2+} deserves careful investigation. Finally, there is the intriguing possibility that zinc ions are normally extruded from presynaptic terminals during transmitter release, as the release we observed is Ca^{2+} -dependent. Until more information is available about

the cellular compartment from which zinc ions are extruded and the mechanisms of this release, the ideas we propose here must remain highly speculative. Whatever the role of transition metal ions in the brain, our findings suggest that Zn^{2+} may be among the endogenous divalent ionic species which modulate neuronal responses.

We thank Brenda Smith, Rosemary Murphy, John Davies and Ruth Slater for technical assistance and Dr M. J. Keating for his helpful suggestions on the manuscript. S.Y.A. is supported by a fellowship from the Wellcome Trust.

Received 21 October 1983; accepted 13 January 1984.

- Owen, C. A. Jr *Wilson's Disease: The Etiology, Clinical Aspects and Treatment of Inherited Copper Toxicosis* (Noyes, Parkridge, New Jersey, 1981).
- Constantinidis, J. & Tissot, R. in *Glutamate as a Neurotransmitter* (eds Di Chiara, G. & Gessa, G. L.) 413–422 (Raven, New York, 1981).
- Chung, S.-H. & Johnson, M. *Expl Neurol.* **82**, 89–107 (1983).
- Pei, Y., Zhav, R., Huang, J. & Cao, L. *Epilepsia* **24**, 169–176 (1983).
- Frederickson, C. J., Klitenick, M. A., Manton, W. I. & Kirkpatrick, J. B. *Brain Res.* **273**, 335–339 (1983).
- Nadler, J. V., Perry, B. W. & Cotman, C. W. *Nature* **271**, 676–677 (1978).
- Le Quesne, P. M. *J. Neurol. Neurosurg. Psychiatr.* **44**, 1–8 (1981).
- Assaf, S. Y., Fan, S., Iversen, L. L. & Lee, C. M. *J. Physiol., Lond.* **308**, 4–5P (1980).
- Chung, S.-H. & Johnson, M. S. *Brain Res.* **280**, 323–334 (1983).
- Frederickson, C. J., Manton, W. I., Frederickson, M. H., Howell, G. A. & Mallory, M. A. *Brain Res.* **246**, 338–341 (1982).
- Evans, R. H. *J. Physiol., Lond.* **298**, 25–35 (1980).
- Ben-Ari, Y., Tremblay, E., Ottersen, O. P. & Naguey, R. *Brain Res.* **165**, 362–365 (1979).
- Smart, T. G. & Constanti, A. *Proc. R. Soc. B215*, 327–341 (1982).
- Smart, T. G. & Constanti, A. *Neurosci. Lett.* **40**, 205–211 (1983).

Stimulation-induced uptake and release of zinc in hippocampal slices

Gailyn A. Howell, Michael G. Welch
& Christopher J. Frederickson

Laboratory for Neurobiology, The University of Texas at Dallas,
Richardson, Texas 75083-0688, USA

The mossy-fibre axons of the hippocampus form a dense plexus, uniquely rich in chelatable zinc^{1–3}. Because the metal is apparently concentrated within the terminal bags of the axons^{4,5}, it has been hypothesized that the zinc is involved in mossy-fibre synaptic transmission^{6,7}. Although some electrophysiological findings^{8,9} have favoured the hypothesis, neither preferential uptake of zinc into the hippocampus^{10,11} nor depolarization-induced release of zinc from hippocampal tissue¹² has previously been found. Using the hippocampal slice preparation, we now report that the mossy-fibre neuropil and cells of origin (dentate granule cells) take up zinc preferentially, and that electrical stimulation selectively facilitates both uptake of exogenous zinc into mossy-fibre neuropil and release of previously incorporated ⁶⁵Zn from the tissue. The results suggest that the role of zinc in mossy-fibre axons is dynamically linked to neural signalling processes.

Our investigation of zinc metabolism in mossy-fibre axons was guided by the finding that only 10% of total hippocampal zinc is associated with mossy fibres¹³. The remaining 90%, representing 'background' zinc (60–80 p.p.m.) distributed throughout all grey matter^{13,14}, presumably reflects the multiple roles of zinc in cell biology¹⁵. We reasoned that even vigorous zinc turnover in mossy fibres would go undetected if studied in the tissue of a whole hippocampal slice. Therefore, to examine mossy-fibre zinc uptake directly, we microdissected hippocampal tissue into cytoarchitectonic fields and compared zinc utilization in the mossy-fibre region with utilization in the other hippocampal regions.

To determine the appropriate conditions for studies of mossy-fibre zinc uptake, we first examined the kinetics of zinc uptake in slices of whole hippocampus. The results indicated that at least two zinc transport systems operate in parallel (Fig. 1a,b). Eadie-Scatchard analysis indicated a high-affinity system ($K_m = 12 \mu M$; $V_{max} = 5.8$ ng per mg per h) and a low-affinity system

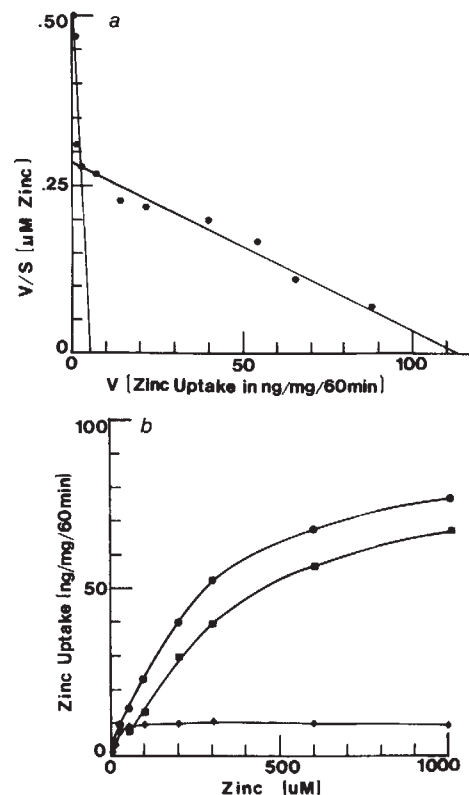


Fig. 1 a, Eadie-Scatchard analysis²⁶ of zinc uptake into hippocampal slices. Velocity of uptake per substrate (0.5–1,000 μM zinc) is plotted as a function of velocity of uptake (ng per mg per 60 min), showing two distinctly separate kinetic processes. Lines of best fit were determined by linear regression analysis ($r > 0.95$) with each point representing the mean for five to seven separate experiments (s.e.m. < 10%). b, Total zinc uptake separated into high-affinity and residual components. The hyperbolic curve A represents the total zinc uptake. Curve B shows uptake by the high-affinity mechanism as demonstrated by the formula $V = V_{max} \times S / (S + K_m)$, where V is the rate of accumulation of zinc in ng per mg per 60 min and S is the μM concentration of zinc in the incubation medium²⁷. The kinetic constants (apparent K_m and V_{max}) were derived from Lineweaver-Burk analysis²⁸ of the data. Curve C shows the mathematical difference between curves A and B (the total uptake minus the high-affinity uptake) and presumably represents the low-affinity uptake. Each point represents the mean for 5–7 experiments (s.e.m. < 10%).

Methods: For the uptake studies, 8–10-week-old Jackson mice (C57BL/6J) were decapitated, the brains rapidly removed, and transverse slices of hippocampus, 0.45 mm thick, were cut mechanically. Slices were preincubated at 37 °C in oxygenated (95% O_2 , 5% CO_2) medium (124 mM NaCl, 5 mM KCl, 26 mM $NaHCO_3$, 10 mM glucose, 2.4 mM $CaCl_2$, 1.24 mM KH_2PO_4 , 1.3 mM $MgSO_4$) for 60 min^{29,30} and were then transferred to a 2-ml well containing a predetermined amount of ⁶⁵Zn diluted with natural zinc to yield a final medium concentration of 0.5–1,000 μM Zn^{2+} . Transport was halted by rinsing the slices in ice-cold substrate-free medium; the slices were then blotted and oven-dried to constant weight. The decay activity of γ rays at 1.11 MeV was measured on a germanium detector. Specific radioactivity of ⁶⁵Zn present in the tissue was compared with that of ⁶⁵Zn in the medium and corrected for efficiency and different specific activities in the medium.

($K_m = 402 \mu M$; $V_{max} = 113$ ng per mg per h). Lineweaver-Burk analysis yielded similar results ($K_m = 18 \mu M$; $V_{max} = 9.2$ ng per mg per h and $K_m = 320 \mu M$; $V_{max} = 103$ ng per mg per h). Subsequent experiments with low concentrations (20 μM) of substrate in the medium revealed that the high-affinity system saturated slowly (over a period of 6 h) and was energy dependent, being markedly suppressed by both cold (42% at 4 °C compared with 37 °C; $P < 0.001$, Student's t -test) and ouabain (54% with 100 μM ouabain in the medium; $P < 0.001$, t -test).

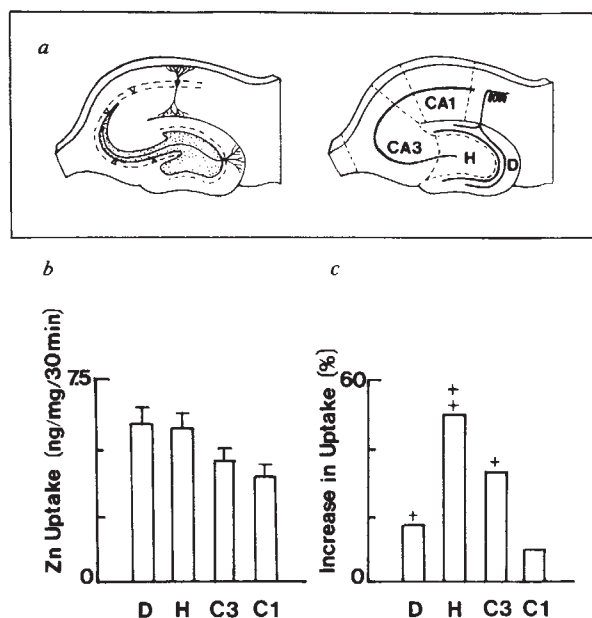


Fig. 2 *a*, The diagram on the left illustrates the pyramidal and granule cell fields of the hippocampus. The stippled area indicates the location of the mossy-fibre neuropil arising from the granule cell layer of the dentate gyrus. The diagram on the right shows the cytoarchitectonic regions of the hippocampus (indicated by dashed lines) that were dissected for assay (the dentate gyrus granule cell/molecular area (D), the hilus of the dentate gyrus (H), and the CA1 and CA3 hippocampal pyramidal cell fields). *b*, Columns represent the baseline uptake velocities for unstimulated tissue (the dentate gyrus granule cell/molecular area (D), the hilus of the dentate gyrus (H), and the CA1 (C1) and CA3 (C3) hippocampal pyramidal cell fields) for nine separate experiments ($\pm$ s.e.m.). Uptake varied significantly among the regions ($P < 0.01$; ANOVA), and uptake into regions D and H was significantly greater ($P < 0.01$; Duncan's test) than uptake into the CA1 region. *c*, Columns represent the percentage increase in uptake ($[(\text{stimulated uptake velocity} - \text{unstimulated velocity}) / \text{unstimulated velocity}] \times 100\%$) produced by electrical stimulation. Stimulation parameters were set at 1.25 times the typical threshold for CA3 evoked field potentials ($0.8 \mu\text{A } 100 \mu\text{m}^{-2}$): 0.4-ms cathodal pulses (capacitatively coupled), $1 \mu\text{A } 100 \mu\text{m}^{-2}$, 25 pulses s^{-1} . Total incubation time in ^{65}Zn was 30 min (including the 25-min period of stimulation). The stimulating electrode (*a*) was made of platinum-iridium wire ($75 \mu\text{m}$ diameter) curved to overlie the granule cell dendrites (uninsulated curved portion was 1.5 mm in length; total current, 2 mA). Mean values from three to four experiments are shown; significance of the stimulated versus unstimulated differences was $P < 0.05$ (+) and $P < 0.001$ (++); *t*-test (s.e.m. $< 10\%$).

Similar zinc uptake kinetics have been observed in other tissues, including striatum and neocortex¹⁶, pancreas¹⁷ and liver¹⁸.

The specific uptake of zinc into the mossy-fibre neuropil was investigated by incubating intact, transverse slices of hippocampus in ^{65}Zn , then dissecting separate cytoarchitectonic regions free for assay. Slices were first incubated in ^{65}Zn -containing medium ($20 \mu\text{M}$; $1.02 \mu\text{Ci ml}^{-1}$) for 30 min. Subsequently, slices were rinsed, blotted, placed flat on glass plates, slowly freeze-dried then dissected into cytoarchitectonic regions (Fig. 2a) under microscopic control (Zeiss, $\times 40$). Tissues from individual regions (pooled from three to four slices) were collected separately on foil, dried to constant weight and assayed for ^{65}Zn activity for γ -ray decay at 1.11 MeV.

The regional analysis revealed a definite relationship between mossy-fibre tissue and zinc utilization (Fig. 2b). Specifically, both the tissue sample (D) comprised of somas and dendrites of the granule cells, and the sample (H) containing the dense mossy-fibre plexus of the hilus incorporated over 40% more zinc than the CA1 sample (C1) containing no mossy fibres. Furthermore, the CA3 sample (C3), containing a single band of mossy fibres, took up an intermediate amount of zinc. Interest-

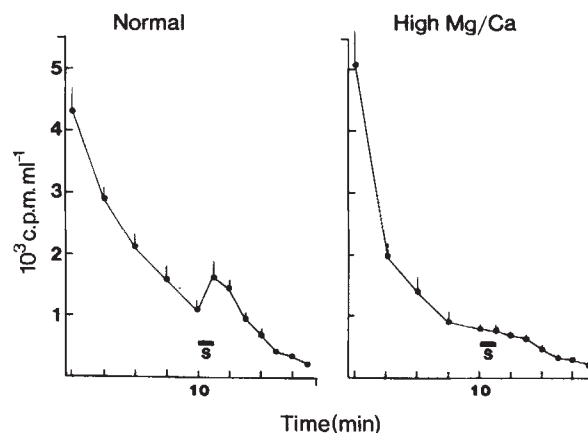


Fig. 3 Radio-zinc efflux into the superfusate before, during and after stimulation applied to the granule molecular zone (electrode placed as in Fig. 2a: $2 \mu\text{A } 100 \mu\text{m}^{-2}$, 12.5 pulses s^{-1} , 0.4-ms cathodal pulses for 60 s; bar marked 'S' shows the collection period corresponding to the time of stimulation treatment. In all release experiments, two hippocampal slices were stimulated simultaneously to double the yield of zinc into the medium. In both panels, means ($\pm$ s.e.m.) from six experiments are shown. The 10-min rinse period immediately preceded the collection of the first aliquot. The superfusion rate averaged 0.75 ml min^{-1} , and the tissue chamber volume was 2.0 ml. Note the absence of stimulation-induced release in high Mg^{2+} (20 mM), low Ca^{2+} (1 mM).

ingly, zinc uptake did not parallel the endogenous zinc content of the tissues. The granule cell region (which incorporated the most zinc) contains considerably less endogenous zinc (80 p.p.m.) than does hilar tissue (150–200 p.p.m.)¹³. It is possible that most of the zinc taken up *in vivo* by granule cells is transported to the mossy-fibre axons by orthograde axoplasmic flow.

Electrical stimulation applied to granule cell dendrites selectively facilitated uptake of zinc into tissue containing mossy fibres (Fig. 2c). The hilar sample showed the largest increase in uptake and CA3 a smaller increase. Uptake into the granule cell region was also slightly enhanced during stimulation, whereas uptake into CA1 tissue was not significantly affected. An additional experiment indicated that the facilitation of zinc uptake was not due to lasting stimulation-induced tissue damage. Slices which were prestimulated in the absence of zinc for 25 min and then subsequently incubated in ^{65}Zn -containing medium ($20 \mu\text{M}$; 30 min) did not incorporate significantly more zinc than did slices treated identically but not stimulated ($P > 0.5$, *t*-test). The enhanced uptake of zinc into the mossy-fibre regions during stimulation of the granule cells may reflect augmented utilization of zinc in neurones, or even in glia. However, because the preponderance of the endogenous zinc in the region is located in mossy-fibre terminals^{4,5}, it is likely that granule cell stimulation facilitated zinc uptake directly into mossy-fibre terminals.

To test for stimulation-induced release of zinc, we first incubated slices of hippocampus in ^{65}Zn -containing medium ($20 \mu\text{M}$; $5.3 \mu\text{Ci ml}^{-1}$; 60 min), then rinsed the slices by a 10-min superfusion with zinc-free medium, and collected serial aliquots of superfusate before, during and after brief electrical stimulation of the granule cells. Stimulation reliably augmented the efflux of ^{65}Zn into the superfusate (Fig. 3). Compared with the 2-min period immediately before stimulation, ^{65}Zn efflux during the 2-min period starting at stimulation onset was increased by 38% (average of six trials; $P < 0.025$, one-tailed *t*-test for matched pre- and post-stimulation samples). Parallel experiments with lowered Ca^{2+} (1 mM) and elevated Mg^{2+} (20 mM) in the superfusion medium indicated that zinc efflux during stimulation was mediated by calcium-dependent mechanisms, consistent with a possible release from presynaptic terminals¹⁹ (Fig. 3).

The present results indicate that the turnover (uptake and release) rate for zinc in mossy-fibre tissue is accelerated by

electrical stimulation of the dentate gyrus. The findings are clearly consistent with the possibility that zinc is involved in synaptic signalling functions of mossy-fibre axons. Whether zinc participates directly in classical synaptic neurotransmission⁶⁻⁹ or is involved in neuromodulator²⁰⁻²² or neurotrophic²³⁻²⁵ messenger functions remains undetermined. Regardless of this, the present data suggest that zinc may be involved in processes dynamically coupled to the electrophysiological activity of the mossy-fibre axons.

Received 21 October 1983; accepted 27 February 1984.

1. Maske, H. *Naturwissenschaften* **42**, 424 (1955).
2. McLardy, T. *Confinia neurol.* **20**, 1-17 (1960).
3. Frederickson, C. J., Howell, G. A. & Frederickson, M. H. *Expl Neurol.* **73**, 812-823 (1981).
4. Haug, F. M. *Histochemie* **8**, 355-368 (1967).
5. Danscher, G. & Zimmer, J. *Histochemistry* **55**, 27-40 (1978).
6. Crawford, I. L. & Connor, J. D. *J. Orthomolec. Psychiat.* **4**, 39-52 (1975).
7. Storm-Mathisen, J. *Prog. Neurobiol.* **8**, 110-181 (1977).
8. Crawford, I. L., Doller, H. J. & Connor, J. D. *Neuropharmacology* **15**, 197 (1973).
9. Hesse, G. W. *Science* **205**, 1005-1007 (1979).
10. Czerniak, P. & Haim, D. B. *Archs Neurol.* **28**, 555-560 (1971).
11. Kasarskis, E. J. *Expl Neurol.* (in the press).
12. Harris, N. F. & Crawford, I. L. *Soc. Neurosci. Abstr.* **7**, 788 (1981).
13. Frederickson, C. J., Kiltienick, M. A., Manton, W. I. & Kirkpatrick, J. B. *Brain Res.* **273**, 335-339 (1983).
14. Donaldson, J., St. Pierre, T., Minnich, J. L. & Barbeau, A. *Can. J. Biochem.* **51**, 87-92 (1973).
15. Vallee, B. L. *Carlsberg Res. Commun.* **45**, 423-441 (1980).

Part of this research was submitted in a doctoral dissertation to the University of Texas at Dallas by G.A.H. This work was supported in part by grants from the NSF (CDP 800879) and the Biomedical Research Support Grant Program, Division of Research Resources, NIH (BRSR-807-RR07133) to C.J.F. Preliminary results were presented at the Society for Neurosciences Meeting, Minneapolis (November 1982) and at the Neurobiology of Zinc Symposium, Boston, Massachusetts (November 1983).

16. Howell, G. A. & Frederickson, C. J. *Soc. Neurosci. Abstr.* **8**, 991 (1982).
17. Ludvigsen, C., McDaniel, M. & Lacy, P. E. *Diabetes* **28**, 570-576 (1979).
18. Failla, M. L. & Cousins, R. J. *Biochim. biophys. Acta* **538**, 435-444 (1978).
19. Rahamimoff, R. in *The Neurosciences: 3rd Study Program* (eds Schmitt, F. O. & Worden, F. G.) 943-952 (MIT Press, Cambridge, 1973).
20. Wu, J. Y. & Roberts, E. *J. Neurochem.* **23**, 759-767 (1974).
21. Ebadi, M., Itoh, M., Bifano, J., Wendt, K. & Earle, A. *Int. J. Biochem.* **13**, 1107-1112 (1981).
22. Stengaard-Pedersen, K. *Acta pharmac. tox.* **50**, 213-220 (1982).
23. Crutcher, K. A. & Davis, J. N. *Soc. Neurosci. Abstr.* **6**, 172 (1980).
24. Stewart, G. R., Frederickson, C. J., Howell, G. A. & Gage, F. H. *Brain Res.* (in the press).
25. Frederickson, C. J. *et al. Neuroscience* (submitted).
26. Segel, I. H. *Biochemical Calculations*, 208-322 (Wiley, New York, 1976).
27. Kiely, M. & Sourkes, T. L. *J. Neurochem.* **19**, 2863-2872 (1972).
28. Lineweaver, H. & Burk, D. *J. Am. chem. Soc.* **56**, 658-666 (1934).
29. Howell, G. A., Brandstetter, J. W., Frederickson, M. H., Frederickson, C. J. & Trulsson, M. E. *Soc. Neurosci. Abstr.* **7**, 583 (1981).
30. Trulsson, M. E., Howell, G. A., Brandstetter, J. W., Frederickson, M. H. & Frederickson, C. J. *Life Sci.* **31**, 785-790 (1982).

Self tolerance is H-2-restricted

Polly Matzinger*, Rose Zamoyska & Herman Waldmann

Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB1 2QP, UK

H-2 restriction is an established characteristic of T-cell behaviour¹⁻³ and, in effect, it means that mouse T cells are activated against foreign antigens only if those antigens are presented in a membrane association with molecules of the mouse major histocompatibility complex, H-2. Whether T-cell inactivation or tolerance is also H-2-restricted is a question which has been tested directly⁴⁻⁵ and indirectly⁶⁻⁷ several times in the past. In each case the answer was 'No' but in each case the answer was inconclusive. Doubts arose because of the observation that activation of T cells, *in vivo*, is an H-2-restricted event which appears unrestricted because of antigen processing by the host⁸⁻⁹. If antigen processing is involved in the induction of tolerance, then tolerance might also be an H-2-restricted process disguised to appear unrestricted. We report here a study designed to minimize antigen processing in which we find that T-cell tolerance induction to 'self' minor histocompatibility (H) antigens is indeed H-2-restricted.

The strains of mice we used carry the H-2^b and/or H-2^d alleles and differ by a variety of minor H antigens. An example of H-2 restriction in T-cell activation is shown in Fig. 1. Normal (B10×B10.D2)F₁ mice (themselves H-2^b×H-2^d) respond well to the foreign minor H antigens carried by BALB/c mice and these responses are H-2-restricted. Those T cells which are activated against BALB/c (H-2^d) lyse BALB/c but not BALB.B (H-2^b) target cells (Fig 1a), and vice versa Fig. 1b. Let us consider whether or not tolerance induction is restricted. (B10×B10.D2)F₁ T cells which have matured in a BALB/c environment should be tolerant of the BALB minor H antigens in association with H-2^d. If tolerance is restricted, such chimaeras should not be tolerant of, and should respond to, these same BALB minor H antigens when presented with H-2^b, a combination of H-2 and antigen to which they have not been exposed. However, if tolerance is not restricted and can be induced by presentation of antigen alone^{5,10}, the chimaeras should be tolerant of the BALB minor H antigens regardless of their associations with H-2.

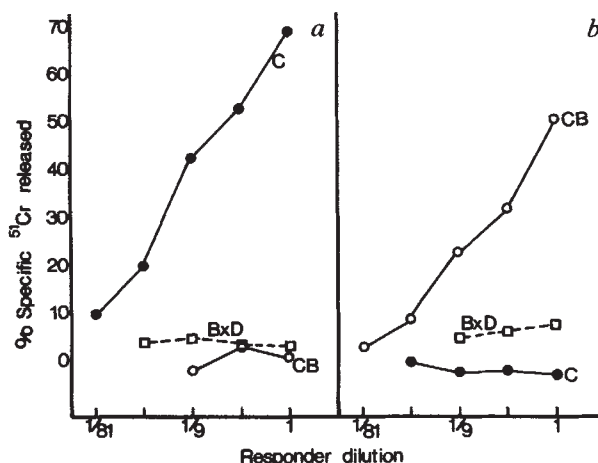


Fig. 1 Mouse strains used in the experiments are given below:

Strain	Abbreviation	H-2 type	Minor H type
B10		b	B10
BALB.B	CB	b	BALB
C3H.SW	SW	b	C3H
B10.D2		d	B10
BALB/c	C	d	BALB
DBA/2	D2	d	DBA
(B10×B10.D2)F ₁	(B×D)	b×d	B10
(BALB.B×BALB/c)F ₁	(CB×C)	b×d	BALB

Cytotoxic response of a normal (B10×B10.D2)F₁ mouse against BALB.B and BALB/c. The mouse was primed at the same time and with the same antigens as the chimaeras (see Fig. 2 legend). Two months after priming, its spleen cells were cultured for 5 days against C (a) or CB (b) irradiated (2,200-3,300 rad) stimulators and assayed for killing activity on 2-3-day concanavalin A (Con A) blasts. Specific ⁵¹Cr release is calculated as (experimental-spontaneous release)/(detergent-spontaneous release). Since effector: target (E:T) ratios vary from mouse to mouse, we present here responder dilution instead. Cells were collected after 5 days in culture, spun down and resuspended to the minimum volume necessary to test on all targets and diluted in threefold dilutions. Highest dilution E:T here is 133:1 calculated from the number of spleen cells originally cultured.

* Present address: Basel Institute for Immunology, Grenzacherstrasse 487, 4005 Basel, Switzerland.

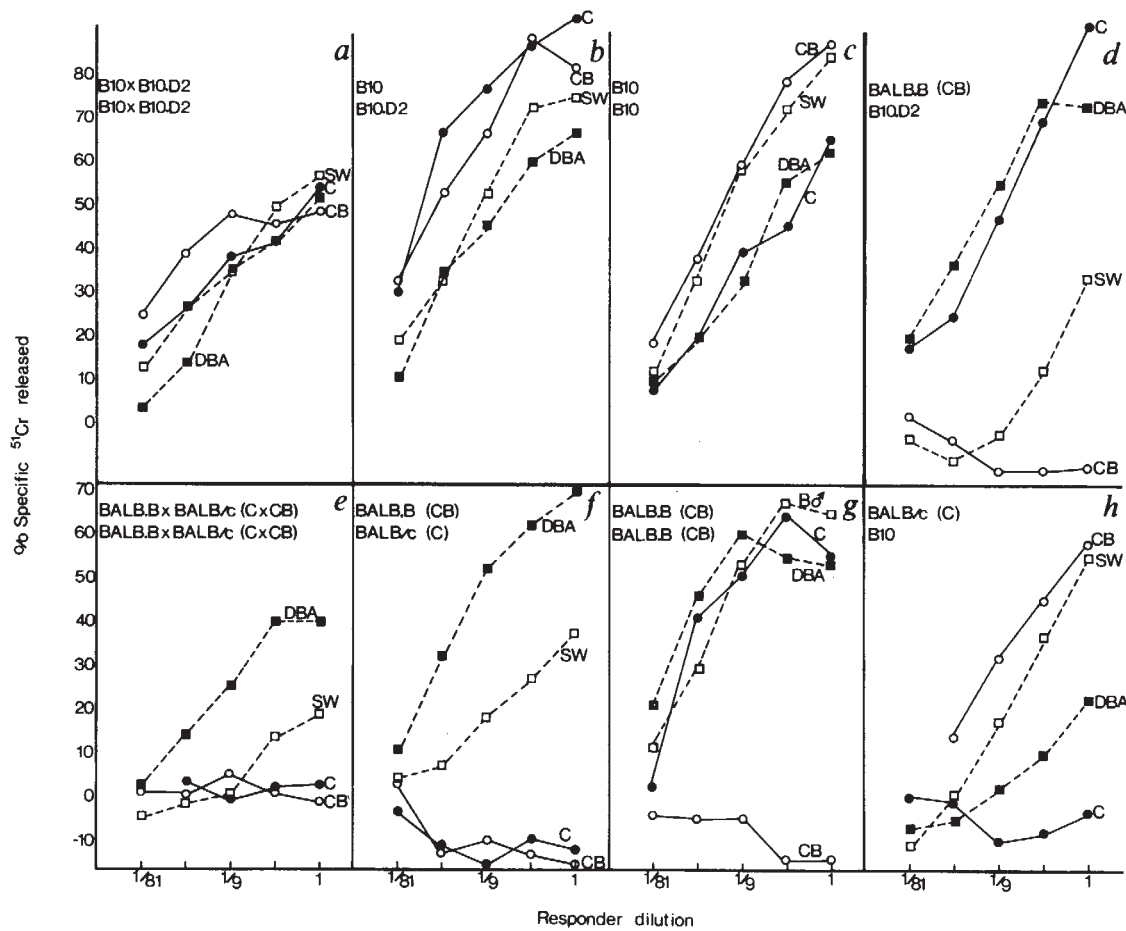


Fig. 2 Cytotoxic responses of thymus-grafted radiation chimaeras to various minor H antigens. The chimaeras were (B10×B10.D2) F_1 mice which had been thymectomized at 6–8 weeks of age, irradiated with 875 rad at 3 months of age and reconstituted 4 h later with anti-Thy 1 + C-treated (B10×B10.D2) F_1 bone marrow. Three weeks later each chimaera was grafted with two irradiated (600 rad) neonatal thymuses under the kidney capsules. Some mice were given two identical thymuses and others were given one each of two different thymuses. Male recipients received male bone marrow and male thymuses, whereas female recipients received female tissues. The combinations described in Table 1 are reported here. Three and a half months after thymus grafting all chimaeras were immunized intraperitoneally (i.p.) with 2×10^6 each of mitomycin-treated (C×CB) F_1 δ and (DBA/2×C3H.SW) F_1 spleen cells. Two to four months later they were given an i.p. boost of a mix of 2×10^6 each C δ , CB δ , DBA/2 δ and C3H.SW δ mitomycin-treated spleen cells. Three weeks to four months later spleen cells from the chimaeras were boosted in 5-day *in vitro* culture and assayed for killing on Con A blasts. Each spleen was split and stimulated separately against C, CB, SW, DBA/2 and/or B10 and B10.D2. Although each culture was tested on several targets, only the killing activity against the stimulating cell type is shown. In no case did any chimaera lyse B10 δ , B10.D2 δ or (B10×B10.D2) F_1 δ targets. Each mouse was ear punched at the time of thymus grafting, and checked at the time of death for the presence of its own intrinsic thymus and its grafted thymuses. Only those mice which carried both grafted thymuses and which had been properly thymectomized are reported. Each box represents an individual mouse, with its grafted thymuses depicted in the upper left corner and the targets against which it was tested indicated on each curve. Highest E:T ratios varied from 89:1 to 133:1 calculated from the number of cells originally cultured.

The main reason why tolerance has always appeared unrestricted may well be antigen-processing; if BALB minor H antigens can be picked up by the (B10×B10.D2) antigen presenting cells and re-presented in association with both H-2^b and H-2^d, then tolerance would always appear unrestricted. We have attempted to minimize the effects of antigen processing by constructing thymus-grafted chimaeric mice carrying the test antigens only on the radioresistant cells of the thymus. Thus the overall dose of antigen is low but the concentration of antigen is high where it may matter most (on the thymic antigen-presenting cells).

Unfortunately, this attempt at solving the processing problem creates another. Over the years, much evidence has accumulated indicating that F_1 cells which have developed in a homozygous parental type thymus may respond poorly or not at all to an antigen when it is presented with the H-2 type of the unshared parent^{11,12}. Therefore, chimaeras grafted with BALB.B (H-2^b) thymuses might not respond to any antigen associated with H-2^d. This problem is not as severe a constraint as it may seem. First, thymic imprinting of T-cell repertoires is not absolute and can often be completely overcome by appropriate immunization^{11,13}. Thus, if chimaeras carrying BALB.B thymuses (H-

2^b) can be immunized to respond against foreign minor H antigens plus H-2^d, they can be tested for tolerance against 'self' minor H antigens plus H-2^d (BALB/c). Second, we grafted some of our chimaeras with two thymuses, only one of which carried the tolerogen. For example, a mouse might be grafted with a BALB.B (H-2^b) thymus (the tolerogen) and a B10.D2 (H-2^d) thymus. The presence of the B10.D2 thymus should enable these mice to respond to foreign minor H antigens in association with H-2^d and allow them to be tested for responses to 'self' antigens plus H-2^d.

We immunized our thymus-grafted chimaeras against a panoply of antigens. Ninety-two male and female chimaeras were immunized and tested for T-cell killing activity against the minor H antigens of BALB/c ('self' + H-2^d), BALB.B ('self' + H-2^b), DBA/2 (foreign + H-2^d) and C3H.SW (foreign + H-2^b), and all the females were immunized against syngeneic male cells. Table 1 shows the predicted responses of mice grafted with different thymuses and Fig. 2 shows the responses of one set of mice, each of which responded to foreign minor H antigens (DBA/2, C3H.SW or δ). Those chimaeras which bore thymuses of the B10 background (a–c) responded to BALB minor H antigens associated with either H-2^d (BALB/c) or H-2^b

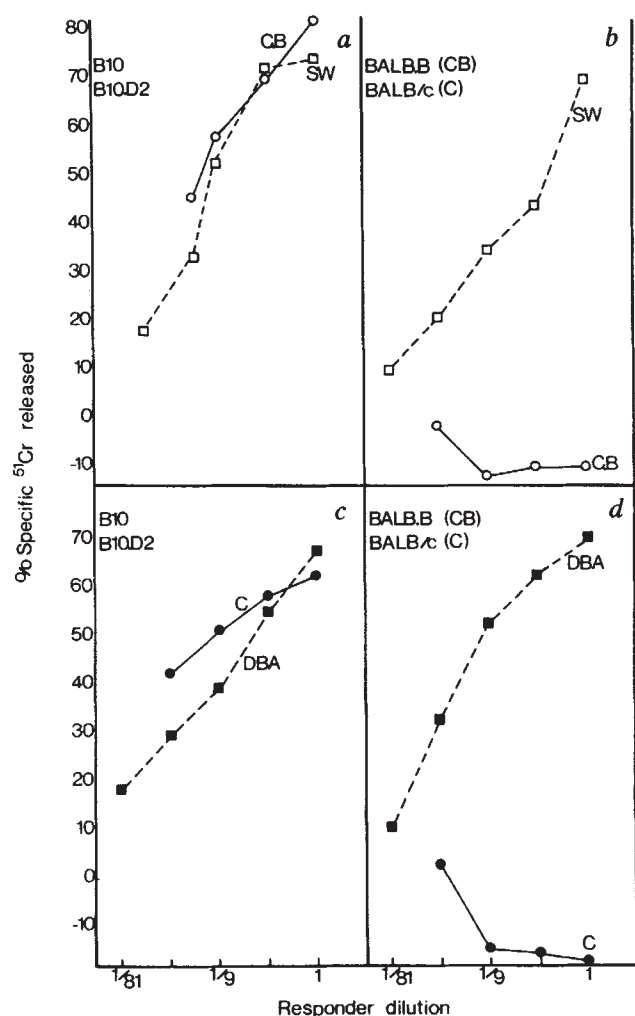


Fig. 3 Cross-reactions of chimaeric killer cells. Spleen cells which had been stimulated in 5-day *in vitro* culture with C3H.SW (a, b) or DBA/2 (c, d) were tested for cytotoxicity on SW and CB or D2 and C targets respectively. Each box represents a single culture of an individual mouse with its grafted thymuses given in the upper left corner. Highest $E:T$ for panel a = 89:1, panels b, c, d = 133:1.

(BALB.B). Chimaeras *e* and *f*, bearing respectively two (BALB/c $\times$ BALB.B) F_1 thymuses or one BALB/c plus one BALB.B thymus were unresponsive to both BALB/c and BALB.B. This, of course, was a necessary condition for the experiment, showing that thymus cells (the organ used to tolerate) express enough antigen to tolerate for spleen cells (the organ used to immunize). Chimaeras *d*, *g* and *h* show that tolerance is H-2-restricted. Chimaera *g*, bearing two BALB.B ($H-2^b$) thymuses, was unresponsive to BALB.B but responded very well to BALB/c. Chimaeras *d* and *h* were reciprocal constructions. Mouse *d* bore a BALB.B thymus (the tolerogen with $H-2^b$) and responded to BALB/c ($H-2^d$), while mouse *h*, bearing a BALB/c thymus, responded to BALB.B.

Is this tolerance only at the level of the helper cell? It has been suggested that helper cells are more susceptible than killer cells to thymic influences^{14,15}, thus the unresponsiveness of a chimaera carrying a BALB.B ($H-2^b$) thymus to BALB.B may be explained not by its lack of killer cells capable of responding but because the killers lack helper cells. The clearest way to test this was to look at killer cell cross-reactions. C3H.SW ($H-2^b$) and BALB.B have many minor H antigens in common (as do BALB/c and DBA/2) as well as several which are different¹⁶. In general, mice of the B10 background will recognize both the shared and unshared antigens so that killer cells raised against C3H.SW will cross-react extensively on BALB.B (Fig. 3a), and killers directed against DBA/2 cross-react on

Table 1 Predicted responses of thymic-grafted mice if tolerance is (is not) H-2-restricted

Thymus under		Target cells	
Left capsule	Right capsule	CB	C
B10 $\times$ B10.D2	B10 $\times$ B10.D2	+(+)	+(+)
B10	B10.D2	+(+)	+(+)
B10	B10	+(+)	+(+)
C $\times$ C.B	C $\times$ C.B	-(-)	-(-)
CB	C	-(-)	-(-)
CB	CB	-(-)	+(+)
B10	C	+(+)	-(-)
B10.D2	CB	-(-)	+(+)

These predictions are based on two fundamental findings: (1) that thymic imprinting (if it exists at all) is not absolute (see Fig. 2c); and (2) that animals bearing one C and one CB thymus are tolerant of both (see Fig. 2f).

BALB/c (Fig. 3c). If the chimaeras carrying BALB.B thymuses are tolerant only at the helper cell level then stimulation with C3H.SW should induce killer cells to the antigens shared with BALB.B, whereas tolerance at the killer cell level should obliterate these killer cells, leaving only those killer cells able to see the unshared antigens on C3H.SW. Figure 3b shows that chimaeras tolerant of BALB.B respond only to the unique antigens on C3H.SW, while chimaeras tolerant of BALB/c respond only to the unique antigens on DBA/2 (Fig. 3d). Therefore, those chimaeras must be tolerant at least at the killer cell level.

Before we can conclude decisively that the tolerance we observed is truly H-2-restricted, we must show that the responses to BALB.B (in animals tolerant of BALB/c) and to BALB/c (in animals tolerant of BALB.B) is actually directed against new associations of the tolerated minor H antigens with the other parental H-2 rather than against any of the unrestricted Qed-1-like antigens which differ between BALB/c and BALB.B (refs 17, 18). We therefore tested the anti-BALB/c killers (from a BALB.B thymus-bearing chimaera) on (BALB.B $\times$ B10.D2) F_1 targets. The T killer cells directed against BALB minor antigens plus $H-2^d$ should not lyse targets carrying the minor H antigens with the wrong H-2 antigens (BALB.B $H-2^b$) or those carrying the appropriate H-2 molecules without the minor H antigen (B10.D2 $H-2^d$). However, they should be active against the F_1 target which has received the right H-2 from one parent and the appropriate minor H antigens from the other. Figure 4b shows that anti-BALB/c killers from a chimaera tolerant of BALB.B behave exactly like cells from a control chimaera (Fig. 4a). They do not lyse BALB.B or B10.D2 targets, but are active against both BALB/c and (BALB.B $\times$ B10.D2) F_1 targets. Figure 4d presents the results for the reciprocal case of a chimaera tolerant of BALB/c which responds to BALB.B. We conclude that the activity is actually directed against the tolerated minor H antigens in a new combination with H-2 molecules and therefore that tolerance is truly H-2-restricted.

H-2-restricted tolerance is the foundation for a model of alloreactivity¹⁹, immune response genes^{9,20} and at least one speculation about autoimmunity²¹. In addition, it may be the selective force against unlimited proliferation of H-2 loci. It has been argued that the polygenism and enormous degree of MHC polymorphism reflects an individual's need to have at least one MHC allele capable of associating with any pathogen. But why not more loci? Most species express a limited number of MHC loci²². An extreme case is seen in frogs where tetraploid and hexaploid species express four (or six) sets of immunoglobulin and haemoglobin genes^{23,24} but have diploidized their MHC genes²⁵. It seems that expression of large numbers of MHC loci is counterproductive. H-2 restriction of tolerance implies that, for every new H-2 locus expressed, an individual must become unresponsive to hundreds of new associations between the new H-2 and all other cell-surface components. The more an individual's repertoire is depleted by self-tolerance, the less

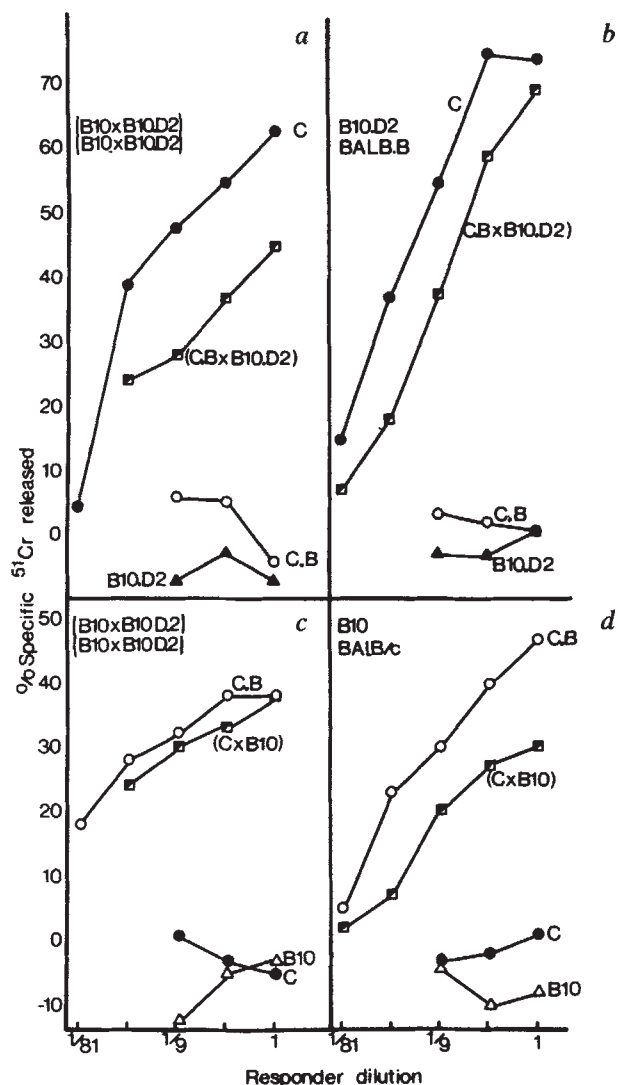


Fig. 4 Chimaeric killer cells are directed against the tolerated minor H antigens in new associations with H-2 molecules. Cells were cultured for 5 days, carried in 20% lectin-free Con A supernatant for 2 days, and restimulated for 3 days before assay on appropriate Con A blasts. Highest $E:T$ panels $a, c = 89:1$, panels $b, d = 22:1$. Panels a, b , stimulated with BALB/c. Panels c, d , stimulated with BALB.B.

there is left over to respond to the world of foreign antigens. Perhaps this is why most animals seem to have a very limited number of MHC loci but the population carries an enormous number of alleles.

That tolerance has always appeared unrestricted in the past must reflect the activity of antigen processing cells. In retrospect this is not surprising. An animal must be tolerant of many antigens not expressed by thymic cells. These might best be presented to young T cells by antigen-presenting cells capable of scouring the body constantly for both self and non-self molecules. We cannot yet say whether these antigen-presenting cells carry such antigens back to the thymus (though they might²⁶), but somehow tolerance must become a dominant activity. Figure 2d, h shows that animals which bear one thymus carrying, and one not carrying, BALB minor H antigens are tolerant. What has happened to the cells which traversed the B10 type thymus? They should respond to the BALB antigens. Are they suppressed²⁷ or have minor H antigens been transported from one thymus to the other?

Since this paper was submitted, Groves and Singer²⁸ have published data which also show that tolerance to minor H antigens is H-2-restricted. In their case the tolerizing antigens were presented by marrow cells in $P \rightarrow F_1$ chimaeras, showing

that the thymus is not unique in its ability to present antigens for MHC-restricted tolerance.

This work was supported by the MRC. P.M. was supported by NIH postdoctoral fellowship AI06052.

Received 8 November 1983; accepted 9 January 1984.

1. Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701-702 (1974).
2. Shearer, G. M. *Eur. J. Immunol.* **4**, 527-533 (1974).
3. Matzinger, P. & Zamoyska, R. *Nature* **297**, 628 (1982).
4. Matzinger, P. & Waterfield, J. D. *Nature* **285**, 492-494 (1980).
5. Lamb, J., Skidmore, B. J., Green, N., Chiller, J. M. & Feldmann, M. *J. exp. Med.* **157**, 1434-1437 (1983).
6. Billingham, R. E. & Silvers, W. E. *J. Immunol.* **85**, 14-26 (1960).
7. Von Boehmer, H., Haas, W. & Helmut, P. *J. exp. Med.* **147**, 1291-1295 (1977).
8. Korngold, R. & Sprent, J. *J. exp. Med.* **151**, 314-327 (1980).
9. Matzinger, P. thesis, Univ. California, San Diego (1979).
10. Cohn, A. & Blomberg, B. *Scand. J. Immunol.* **4**, 1-24 (1975).
11. Fink, P. J. & Bevan, M. J. *J. exp. Med.* **148**, 766-775 (1978).
12. Zinkernagel, R. M. *et al. J. exp. Med.* **147**, 882-896 (1978).
13. Matzinger, P. & Mirkwood, G. *J. exp. Med.* **148**, 84-92 (1978).
14. Singer, A., Hathcock, K. S. & Hodes, R. *J. exp. Med.* **155**, 339-344 (1982).
15. Kruisbeek, A. M., Sharrow, S. O., Mathieson, B. J. & Singer, A. *J. Immunol.* **127**, 2168-2176 (1981).
16. Klein, J. *Transplantation* **15**, 137-153 (1973).
17. Flaherty, L. *Immunogenetics* **3**, 533-539 (1976).
18. Lindahl, K. *Immunogenetics* **8**, 71-76 (1979).
19. Matzinger, P. & Bevan, M. J. *Cell Immunol.* **29**, 1-5 (1977).
20. Schwartz, R. H. *Scand. J. Immunol.* **7**, 3-10 (1978).
21. Matzinger, P. *Nature* **292**, 497-501 (1981).
22. Gotze, D. (ed.) *The Major Histocompatibility System in Man and Animals* (Springer, Berlin, 1977).
23. Du Pasquier, L. & Blomberg, B. *Immunogenetics* **15**, 251-260 (1982).
24. Jeffreys, A. J. *et al. Cell* **21**, 555-564 (1980).
25. Du Pasquier, L. *et al. Immunogenetics* **5**, 129-141 (1977).
26. Fink, P. J., Bevan, M. J. & Weissman, I. L. *J. exp. Med.* (in the press).
27. Morrissey, P. J., Bradley, D., Sharrow, S. O. & Singer, A. *et al. J. exp. Med.* **158**, 365-377 (1983).
28. Groves, E. & Singer, A. *J. exp. Med.* **158**, 1483-1497 (1983).

Evidence from *in vitro* studies that tolerance to self antigens is MHC-restricted

Hans-Georg Rammensee & Michael J. Bevan

Department of Immunology, Scripps Clinic and Research Foundation, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

Mature T cells respond to foreign antigens in the context of self major histocompatibility complex (MHC)-encoded products: T helper cells recognize antigen in the context of class II^{1,2} molecules, while cytotoxic T cells (CTL) recognize antigen plus class I molecules³⁻⁶. Recent evidence suggests that the MHC-restricted T cell is unable to recognize either the foreign antigen or the self-MHC product alone, but only a complex of the two⁷⁻⁹. Unresponsiveness to self antigens—self tolerance—implies the deletion or suppression of clones of T cells having reactivity to self antigens^{10,11}. Here we demonstrate the presence in normal mice of T cells which recognize self antigens together with allogeneic MHC products. This finding suggests the MHC restriction of T-cell recognition during the entire process of T-cell ontogeny, that is, MHC restriction of self tolerance.

We wished to test for the existence in normal mice of allorestricted CTL having specificity for self minor histocompatibility antigens. Minor histocompatibility antigens are cell surface antigens of unknown function encoded in at least 40 loci scattered throughout the mouse genome. To detect allorestricted T cells, one first has to remove T cells which recognize allogeneic H-2 molecules alone. To do this we chose the *in vitro* technique of bromodeoxyuridine (BrdUrd) and light-induced suicide of proliferating cells¹². This method has been used successfully for the detection of allorestricted T helper cell responses to soluble antigens¹³. Allorestricted CTL, however, have not been obtained so far by the BrdUrd and light method¹⁴, although hapten-specific, or minor histocompatibility-specific allorestricted CTL responses were readily demonstrated by the other methods of removing alloreactive cells¹⁵⁻¹⁸. By using the CTL response to the hapten trinitrophenol (TNP) as a prototype³,

Fig. 1 C57BL/10 (B10, H-2^b) spleen cells were stimulated with irradiated (3,000R) B10.BR (H-2^k) spleen cells at a ratio of 1:1 at 4×10^6 nucleated cells ml⁻¹ in RPMI 1640 culture medium, supplemented with 10% fetal calf serum, in a volume of 30 ml in 250-ml culture flasks (Co-star no. 3075) in an upright position at 37 °C, 5% CO₂. On the next day (day 1), 5-bromo-2'-deoxyuridine (BrdUrd) (Sigma) was added. The BrdUrd concentration used was 10^{-6} M (a), 3×10^{-6} M (b), 10^{-5} M (c) or 3×10^{-5} M (d). On days 3 and 4, the cultures were exposed to light¹⁴ (a 40 W fluorescent light tube; Syl- vania F40CW) for 90 min at a distance of 10 cm, the flasks being held in a horizontal position. After the last illumination, the surviving responder cells were collected, washed once, then restimulated with TNP-modified (10 mM trinitrobenzenesulphonic acid for 10 min at 37 °C), irradiated B10.BR spleen cells in the presence of 5% rat Con A supernatant and 50 mM α -methyl-D-mannoside. After another 4 days, the responder cells were collected and assayed on B10 (○), B10.BR (□) and TNP-modified B10.BR (△) Con A blasts in a ⁵¹Cr-release assay as described elsewhere²⁸. Spontaneous ⁵¹Cr-release was between 22.8% and 26.7%. R:T, responder/target ratio.

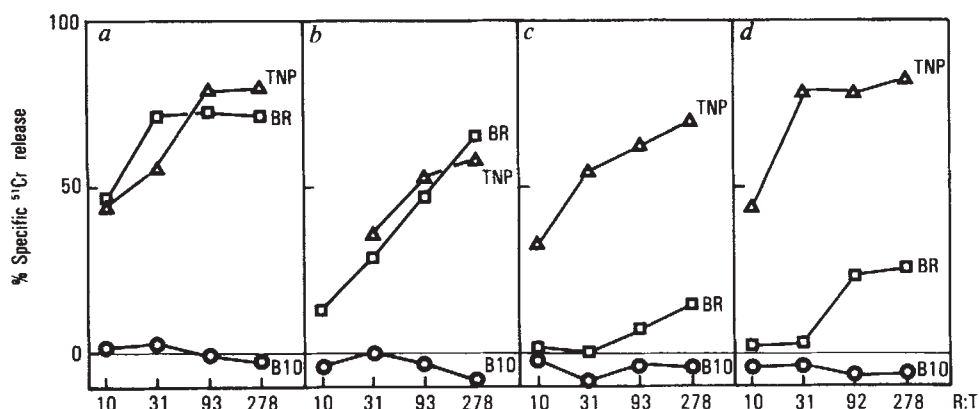
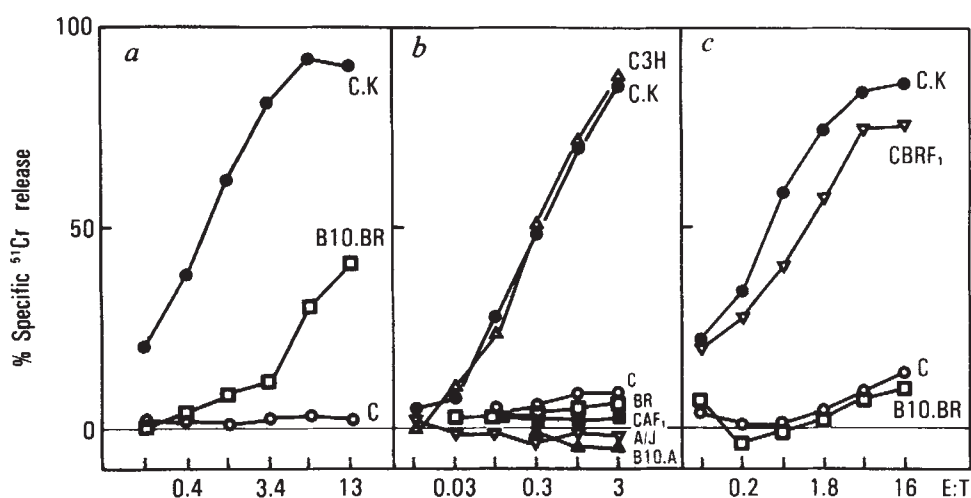


Fig. 2 BALB/c (H-2^d) spleen cells were stimulated with B10.BR spleen cells in the conditions described in Fig. 1 legend. BrdUrd (6×10^{-5} M) was added on day 1 and on days 3, 4, 5, 6 and 7 the cultures were exposed to light for 90 min. Fresh BrdUrd (10^{-4} M), 5% rat Con A supernatant and 50 mM α -methyl-D-mannoside were added at day 4. The surviving cells were primed *in vitro* with irradiated BALB.K (H-2^k) cells in the presence of rat Con A supernatant and α -methyl-D-mannoside for 7 days, then restimulated in the same conditions. a, On day 4 of the restimulation (day 18 of total culture) a sample of the culture cells was tested in a ⁵¹Cr-release assay on Con A blasts of BALB.K (●), B10.BR (□) and BALB/c (○). The responder cells were restimulated every 7 days; ⁵¹Cr-release assays were done 4 days after restimulation. b, After the third restimulation, C3H/St (△), BALB.K (●), BALB/c (○), B10.BR (□), A/J (▽), B10.A (▲) and (BALB/c × A/J)_{F1} (■) targets were tested. After the ninth restimulation BALB/c (○), BALB.K (●), B10.BR (□) and (BALB/c × B10.BR)_{F1} (▽) targets were tested. The spontaneous ⁵¹Cr-release of the individual targets was between 16.5% and 29.9%. E:T, effector/target ratio.



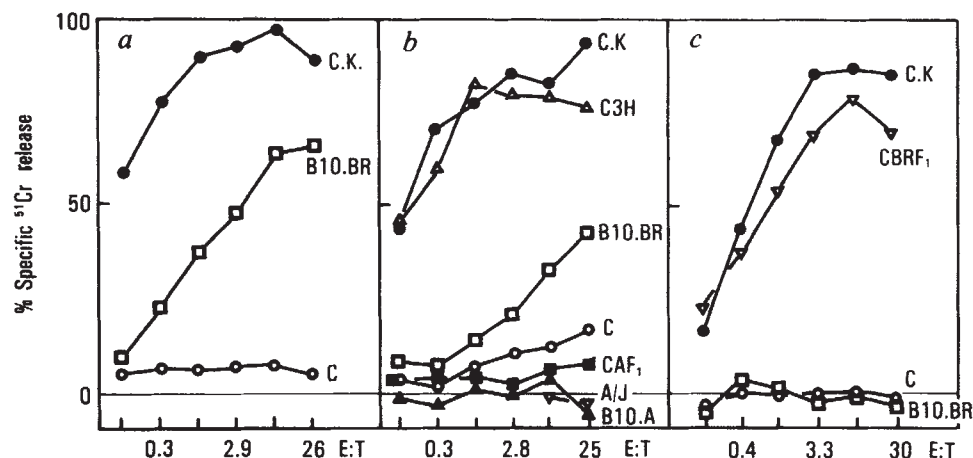
we devised a protocol which made it possible to obtain allorestricted CTL by the BrdUrd and light method.

C57BL/10 (B10, H-2^b) spleen cells were acutely depleted of B10.BR (H-2^k) reactive T cells during a 4-day mixed lymphocyte culture using BrdUrd and light exposure and restimulated for another 4 days with irradiated TNP-modified B10.BR spleen cells (Fig. 1). Insufficient concentrations of BrdUrd during the depletion culture resulted in a CTL population which could lyse unmodified B10.BR targets to the same extent as TNP-B10.BR targets. Higher concentrations of BrdUrd, however, removed most of the B10.BR reactive cells, leaving the TNP-B10.BR reactive CTL. The low remaining alloreactivity is in agreement with the earlier finding that BrdUrd plus light can remove most (up to 99%), but not all, of the alloreactive CTL¹⁹. Attempts to use the same protocol for the detection of minor histocompatibility antigen-specific, allorestricted CTL proved to be unsuccessful, probably due to a lower precursor frequency of allorestricted, minor histocompatibility-specific CTL compared with TNP-specific CTL, resulting in a depleted culture in which the remaining alloreactive CTL precursors outnumbered the allorestricted ones. Therefore, we modified the protocol and used a more stringent procedure for the depletion, lasting 7 days and exposing the cultures on 5 subsequent days to light; in addition, interleukin 2 was added at day 4 to force more CTL precursors specific for allo-MHC into proliferation.

Cultures of BALB/c (H-2^d) spleen cells depleted in this way of reactivity against B10.BR (H-2^k) stimulator cells, lost 97%

of the input of viable nucleated cells. The remaining cells were stimulated twice *in vitro* with irradiated BALB.K (H-2^k) spleen cells, which gave a CTL population showing strong killing of BALB.K, but only a low level of lysis of B10.BR concanavalin A (Con A)-stimulated blasts (Fig. 2a). Repeated restimulations at 7-day intervals resulted in a loss of the remaining B10.BR reactivity, leaving a CTL line (called 25-1) that was presumably specific for BALB minor histocompatibility antigens in the context of H-2D^k, as shown by the reactivity pattern of this line (Fig. 2b,c). BALB.K, (BALB/c × B10.BR)_{F1} and C3H (H-2^k) targets were lysed, but not BALB/c, B10.BR, A/J (K^dD^d), B10.A (K^dD^d) or (BALB/c × A/J)_{F1}. C57BL/10 (H-2^b) and BALB.B (H-2^b) targets, tested in a separate experiment (data not shown), were not lysed. The observed complementation of two parental genes in an F₁ cell, namely the lysis of (BALB/c × B10.BR)_{F1} Con A blasts, but not BALB/c or B10.BR homozygotes, formally proves the specificity of this CTL line as an H-2-restricted one: the (BALB/c × B10.BR)_{F1} cells express the H-2^k products of B10.BR together with BALB minor histocompatibility products on their surface. This F₁ complementation probably excludes the involvement of a Qa/Tla region-encoded antigen²⁰. The recognized BALB minor histocompatibility antigen is H-2D^k-restricted, as shown by the lack of killing of (BALB/c × A/J)_{F1} targets. The background gene is also expressed on C3H cells, which is not surprising considering the low degree of polymorphism of minor histocompatibility genes²¹. The 25-1 CTL line also lyses C3H.OH (K^dD^k) targets, supporting the mapping to H-2D^k.

Fig. 3 This is an independent experiment done using the same strain combination and in the same conditions as in Fig. 2, but with one difference: the depletion culture was terminated and the *in vitro* priming started on day 6 of the culture. Assays were done after the first (a), third (b) and ninth (c) restimulations.



A second, independent experiment (Fig. 3) was done in the same strain combination with one slight modification of the depletion protocol—the culture time was shortened by 1 day, which resulted in a higher number of surviving cells (7%), but also in a higher remaining alloreactivity (Fig. 3a). This alloreactivity again faded out on repeated restimulations with BALB.K. The reactivity pattern of the resulting CTL line, 26-5, was identical to that of line 25-1 (Fig. 3b,c). In addition, with line 26-5 the remaining alloreactivity was entirely H-2D^k-specific, because B10.BR (K^kD^k) were killed, but B10.A and A/J (K^dD^d) were not. This lack of K^k alloreactivity parallels the absence of K^k + BALB minor histocompatibility-specific CTL in our lines.

Thus, both independently derived CTL lines, 25-1 and 26-5, recognize a BALB minor histocompatibility antigen together with an H-2D^k product. Because the CTL originated from BALB/c mice, they recognize a self minor histocompatibility antigen in the context of a foreign H-2 class I product.

Similar results were obtained in a different strain combination: B10 spleen cells were depleted of BALB/c reactivity by the BrdUrd and light method and stimulated with B10.D2 (H-2^d) in limiting dilution conditions, revealing B10.D2-specific clones which did not lyse BALB/c.

Our data show the existence of CTL in the spleens of normal mice which are capable of attacking self cell membrane antigens, provided they are recognized in the context of foreign MHC products. This illustrates that during the ontogeny of the CTL repertoire, clonal deletion of self-reactive cells occurs only via MHC-restricted T-cell recognition, that is, the induction of self tolerance is MHC-restricted. However, the following objections may be raised against this interpretation.

(1) The nominal antigen recognized by the CTL lines 25-1 and 26-5 is encoded in the BALB genome, but only expressed on the cell surface when H-2^k genes are also expressed. Such H-2 allele-specific regulation of non-H-2 genes has not been previously reported. Alternatively, the BALB non-MHC gene may regulate the expression of a class I-like antigen present in the MHC of BALB.K. This would resemble the situation described by Fischer Lindahl *et al.*²² for a maternally controlled factor.

(2) The antigen binding receptors of our CTL lines were constructed during the *in vitro* culture by somatic mutation from H-2^k-reactive precursors. We feel this to be an unlikely possibility, although we cannot exclude it.

The MHC restriction of self tolerance is suggested also by a set of data obtained recently in a completely different system²³, in which minor histoincompatible [parent → F₁] chimaeras were shown to contain CTL precursors specific for parental minor histocompatibility antigens + F₁ MHC, although these chimaeras were tolerant to parental as well as F₁ cells. These data agree well with the results of our *in vitro* system. Our findings are paralleled by the recent demonstration of allorestricted, but not self-restricted, T helper cells in guinea pigs capable of recognizing self insulin determinants²⁴. Collectively, these findings support the clonal deletion model for the immune response gene

control^{25,26}. According to this hypothesis, some strains may be non-responders to certain foreign antigens (seen in conjunction with self-MHC) because they have been tolerized by exposure to a self antigen also seen in the context of self-MHC. The observation that T helper cells from a non-responder strain can recognize antigen on antigen-presenting cells from another non-responder strain, if they differ only in their MHC²⁷, is extended by our data to the CTL level because BALB/c, as well as BALB.K, can be considered as non-responders to BALB minor antigens as long as the latter are presented together with syngeneic MHC products.

A second MHC-restricted level of self tolerance is suggested by recent findings from several laboratories demonstrating the existence of a special T-T cell interaction which results in the suppression of a CTL response²⁸⁻³⁰. CTL precursors recognizing a minor histocompatibility antigen together with a class I product³¹, or a class I product alone (our unpublished observation) are prevented from functioning as CTL by other T cells expressing these antigens on the surface. The latter have been called 'veto' cells³². From the point of view of these veto cells, CTL responses against self antigens are suppressed. As these cells show strong, specific suppressor activity *in vivo*, they may well have a role in maintaining self tolerance.

In addition to our main conclusion, these data suggest that the generation of allorestricted T-cell responses could be useful for the demonstration of antigens usually not detected in syngeneic situations. Such antigens could include tumour-specific antigens on nonimmunogenic tumours, or tissue-specific, monomorphic antigens for which all members of a species are non-responders, as long as these antigens are presented in the context of self-MHC products.

This work was supported by USPHS grants AI 19335 and AI 19499. H.-G.R. is the recipient of grant Ra 369/1-1 of the Deutsche Forschungsgemeinschaft. We thank Julie McLain for secretarial help and Dr Pam Fink for reading the manuscript.

Received 9 December 1983; accepted 6 February 1984.

- Katz, D. H., Hamaoka, T. & Benacerraf, B. *J. exp. Med.* **137**, 1405-1418 (1973).
- Rosenthal, A. J. & Shevach, E. *J. exp. Med.* **138**, 1194-1418 (1973).
- Shearer, G. M. *Eur. J. Immun.* **4**, 527-533 (1974).
- Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701-702 (1974).
- Bevan, M. J. *Nature* **256**, 419-421 (1975).
- Gordon, R. D., Simpson, E. & Samuelson, L. E. *J. exp. Med.* **142**, 1108-1120 (1975).
- Kappler, J. W. *et al. J. exp. Med.* **153**, 1198-1214 (1981).
- Hunig, T. & Bevan, M. J. *Nature* **294**, 460-462 (1981).
- Heber-Katz, E. *et al. J. exp. Med.* **155**, 1086-1099 (1982).
- Nossal, G. J. V. & Pike, B. L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3844-3847 (1981).
- Dorsch, S. & Roser, B. J. *J. exp. Med.* **145**, 1144-1157 (1977).
- Zoschke, D. C. & Bach, F. H. *Science* **170**, 1404-1406 (1970).
- Ishii, N. *et al. J. exp. Med.* **154**, 978-982 (1981).
- Janeway, C. A. *et al. J. exp. Med.* **147**, 1065-1077 (1978).
- Wilson, D. B. *et al. J. exp. Med.* **146**, 361-367 (1977).
- Forman, J. & Streilein, J. W. *J. exp. Med.* **150**, 1001-1007 (1979).
- Stockinger, H. *et al. J. exp. Med.* **153**, 1629-1639 (1981).
- Teh, H.-S., Bennink, J. & Von Boehmer, H. *Eur. J. Immun.* **12**, 887-892 (1982).
- Maryanski, J. L., Cerottini, J.-C. & Brunner, K. T. *J. Immun.* **124**, 839-845 (1980).
- Lindahl, K. *Immunogenetics* **8**, 71-76 (1979).
- Rammensee, H.-G. & Klein, J. *Immunogenetics* **17**, 637-647 (1983).
- Fischer Lindahl, K., Hausmann, B. & Chapman, V. M. *Nature* **306**, 383-385 (1983).

23. Groves, E. S. & Singer, A. *J. exp. Med.* **158**, 1483-1497 (1983).
24. Dos Reis, G. A. & Shevach, E. M. *J. exp. Med.* **157**, 1287-1299 (1983).
25. Schwartz, R. H. *Scand. J. Immun.* **7**, 3-10 (1978).
26. Matzinger, P. *Nature* **292**, 497-501 (1981).
27. Ishii, N., Nagy, Z. A. & Klein, J. *Nature* **295**, 531-533 (1982).
28. Fink, P., Weissman, I. L. & Bevan, M. J. *J. exp. Med.* **157**, 141-154 (1983).
29. Miller, R. G. *Nature* **287**, 544-546 (1980).
30. Rammensee, H.-G., Nagy, Z. & Klein, J. *Eur. J. Immun.* **12**, 930-934 (1982).
31. Rammensee, H.-G., Juretic, A., Nagy, Z. A. & Klein, J. *J. Immun.* **132**, 668-672 (1984).
32. Muraoka, S. & Miller, R. G. *J. Immun.* **131**, 45-49 (1983).

An interaction between vinculin and talin

Keith Burridge & Paul Mangeat

Laboratories for Cell Biology, Department of Anatomy, The School of Medicine, The University of North Carolina at Chapel Hill, North Carolina 27514, USA

In cultured fibroblasts, microfilament bundles terminate at adhesion plaques (focal contacts), the specialized regions where the cells adhere most tightly to the underlying substrate. Vinculin is a protein concentrated in adhesion plaques¹ and has been suggested as a possible link between the ends of the bundles of actin filaments and the plasma membrane^{1,2}. If vinculin is one protein in a chain of attachment between the bundles of microfilaments and the plasma membrane, it is important to identify other components which interact with vinculin. We have recently discovered a new protein in adhesion plaques which we refer to as talin^{3,4}. Here we show that talin binds to vinculin, which suggests that talin may be involved with vinculin in the attachment of microfilament bundles to the plasma membrane at the adhesion plaques.

Evidence for an interaction between talin and vinculin came initially from SDS-polyacrylamide gel analysis of sucrose density gradient centrifugation experiments (Fig. 1). When vinculin and talin are sedimented through separate gradients the peak concentrations of the proteins have sedimentation coefficients of approximately 6S and 8S, respectively. On sedimentation of a mixture of the proteins, these peak values are essentially unchanged, but a significant fraction of the proteins shifts to a lower (more dense) position in the gradient. When talin is centrifuged alone, it is first detected in the gradient in fraction 14; similarly, when vinculin is centrifuged alone it is first detected in fraction 17. When the two proteins are mixed (Fig. 1, third panel) they are both detected in fraction 12, which represents a shift in apparent sedimentation coefficient of from 7.6 to 10.5 for vinculin and from 9.4 to 10.5 for talin. Only a fraction of the vinculin moved into this complex with talin, but it is interesting that a proteolytic degradation product of vinculin (J. R. Feramisco and K.B., unpublished observation) with a molecular weight (M_r) of 100,000 (just visible in the gel of this preparation) bound more efficiently to the talin. In this case the binding appeared to be almost complete and the sedimentation coefficient for the vinculin fragment shifted dramatically from about 5.3S to 9.4S.

The binding of vinculin to talin has also been confirmed in a gel overlay assay (Fig. 2). This approach was modelled on recent experiments by Otto⁵ in which radioiodinated vinculin was used in gel overlays to identify vinculin-binding proteins. These experiments revealed a high-molecular weight vinculin-binding protein in gels of gizzard and fibroblast extracts⁵. By using radioiodinated vinculin we have demonstrated that this high-molecular weight component is indeed talin (Fig. 2a) and have also identified several other potential vinculin-binding proteins in gels of crude gizzard extracts (Fig. 2a, lanes 1, 1'). These proteins have molecular weights of 200,000 (200K), 180K and 100K. The 200K component co-migrates with myosin but no binding is found to the myosin band of chick embryo fibroblasts (Fig. 2a, lanes 2, 2'), suggesting that this component may be

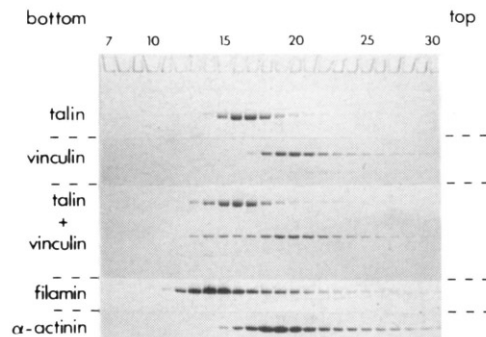


Fig. 1 Section of SDS-polyacrylamide gels of fractions from four sucrose density gradients on which were sedimented talin alone, vinculin alone, talin and vinculin mixed together, or α -actinin and filamin together. When talin is centrifuged alone (top panel) it is first detected in fraction 14 of the sucrose gradient, the peak concentration occurring in fractions 16 and 17. Similarly, when vinculin is sedimented alone (second panel) it is first detected in fraction 17 and the peak occurs in fraction 20. When the two proteins are mixed (third panel) the peaks of the two proteins are essentially unchanged but now the first fraction in which either protein can be detected is 12, representing a shift in apparent sedimentation coefficient of from 7.6S to 10.5S for vinculin and from 9.4S to 10.5S for talin, suggesting a complex between these two proteins. α -Actinin (6.4S) and filamin (9.0S) were centrifuged on a parallel sucrose density gradient as control proteins (bottom panels). Note that there is a 100K degradation product of vinculin (J. R. Feramisco and K.B., unpublished observations) that binds quantitatively to the talin as evidenced by a shift in its position in the gradients from fractions 20-24 to fractions 13-16 when mixed with talin.

Methods: Vinculin was purified as described previously¹⁴ with an additional step of gel filtration through Sepharose C1-6B. Talin was purified by a modification of our previous procedure³, omitting the initial phosphocellulose column. α -Actinin and filamin were purified as described previously¹⁴. Before sedimentation the proteins were incubated separately or mixed together for 7 h on ice in 150 mM NaCl, 0.1% NaN₃, 0.1% β -mercaptoethanol, 50 mM Tris-HCl pH 7.5, then layered over linear 5-20% sucrose density gradients prepared in the same buffer. The gradients were centrifuged at 4 °C for 20 h in a Beckman SW41 rotor at 40,000 r.p.m. Each gradient was collected as 36 $\times$ 11 drop fractions and fractions 7-30 were analysed on 10% polyacrylamide gels¹⁵. Photographs of sections of the gels containing the protein bands are shown aligned according to fractions for easy comparison of the position in the gradients to which the proteins had sedimented.

distinct from myosin. The 100K protein co-migrates with α -actinin, but further work will be needed to confirm its identity. The reciprocal experiment in which radioiodinated talin was used to overlay gels, again revealed that vinculin is the major talin-binding protein in crude cell extracts as assayed by this method (Fig. 2b). In cells of gizzard extracts a second prominent band with a molecular weight of 150,000 is detected (Fig. 2b, lanes 1, 1'). This polypeptide co-migrates with a gizzard protein that is structurally and immunologically related to vinculin^{6,7}. In gels of chick embryo fibroblasts, where this protein is absent⁶, no binding is detected at this molecular weight (Fig. 2b, lanes 2, 2').

The binding of vinculin to talin has been quantitated using a specific immunoprecipitation assay (Fig. 3). Constant amounts of talin were immunoprecipitated after incubation with ¹²⁵I-labelled vinculin. The specificity of the assay was demonstrated by the following criteria: only native iodinated vinculin bound to the talin (heat-denatured ¹²⁵I-vinculin failed to bind, data not shown); addition of increasing concentrations of unlabelled vinculin displaced the binding of ¹²⁵I-vinculin (Fig. 3), whereas an excess of α -actinin instead of vinculin had no effect (data not shown). The binding data were subjected to Scatchard analysis⁸ (Fig. 3, inset) which indicates a high-affinity interaction between the two proteins with an extrapolated, apparent K_d of 10^{-8} M and with ~ 0.3 molecule of vinculin binding per talin

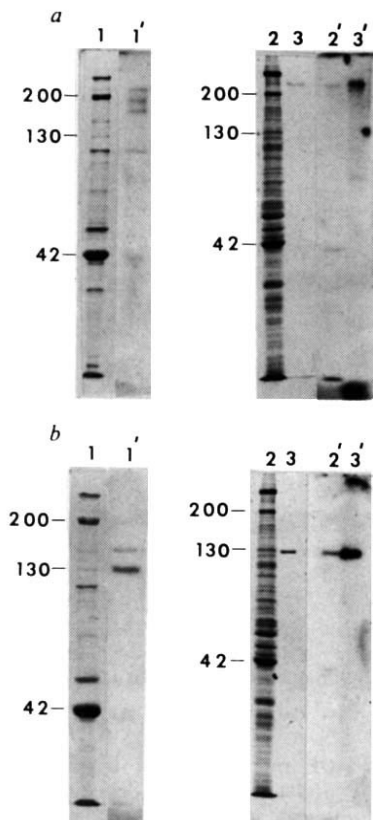


Fig. 2 Direct detection of vinculin-binding (*a*) and talin-binding (*b*) proteins in SDS gels. Photographs of Coomassie blue-stained gels and the corresponding autoradiographs are shown. In *a*, the gel slices correspond to a low ionic strength extract of chicken gizzard (lane 1), whole chicken embryo fibroblasts (lane 2), and purified chicken gizzard talin (lane 3). The autoradiographs of parallel gel slices overlaid with ^{125}I -vinculin are shown (lanes 1', 2', 3'). Note that the major vinculin-binding protein in the gel of whole fibroblasts corresponds to talin (*a*, lane 2'), whereas in the gel of the gizzard extract, bands that bind vinculin are also seen with apparent molecular weights of 200K, 180K and 100K (*a*, lane 1'). Some binding to the actin band is seen in both autoradiographs and this may reflect the affinity of vinculin for actin or a nonspecific interaction, since many iodinated proteins when denatured bind nonspecifically to the actin band in gel overlays¹⁶. In *b*, gels of the crude chicken gizzard extract (lane 1), chicken embryo fibroblasts (2) and pure chicken gizzard vinculin (3), are shown, together with parallel autoradiographs (lanes 1', 2', 3') of gels that were overlaid with ^{125}I -talin. The most prominent talin-binding component corresponds to the vinculin band at M_r 130,000 in gels of both the gizzard extract (*b*, lane 1') and whole fibroblasts (*b*, lane 2'). A second major band, with an apparent molecular weight of $\sim 150,000$, is seen in the autoradiograph of the gizzard extract (*b*, lane 1'). Vinculin and talin were radioiodinated using the chloramine-T method¹⁷ and ^{125}I from NEN. The iodinated proteins were obtained with specific activities of 7×10^6 c.p.m. μg^{-1} for vinculin and 10^7 c.p.m. μg^{-1} for talin. These were layered over gels at 2×10^7 c.p.m. ml^{-1} for 2 h. Gels were prepared for the overlay procedure as described previously^{16,18}. The gel slices in this figure were not from the same slab gel and the position of migration of the bands differs slightly between the gels. The positions of molecular weight marker proteins ($M_r \times 10^{-3}$) are indicated next to each gel. The marker proteins were myosin (200K), vinculin (130K) and actin (42K).

molecule. The Scatchard plot deviates from a straight line, indicating lower affinity interactions. Extrapolation from the curve (not shown) suggests a second binding site with a K_d of $\sim 2 \times 10^{-8}$ M.

A sequence of proteins has been identified linking actin to the erythrocyte membrane (reviewed in ref. 9); similarly, a set of proteins may be involved in membrane attachment of actin at the adhesion plaque of cultured fibroblasts. The interaction

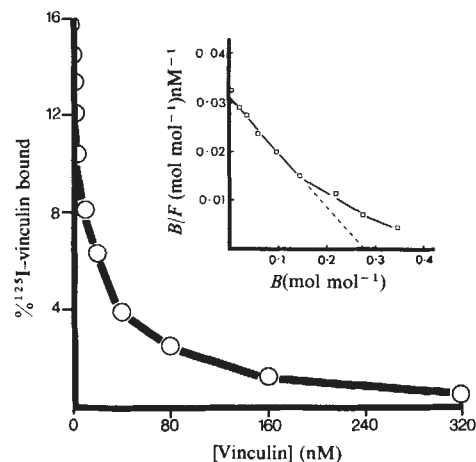


Fig. 3 Binding of ^{125}I -vinculin to talin. Talin ($0.25 \mu\text{g}$) was incubated for 12 h at 0°C in a 0.2-ml volume containing ^{125}I -vinculin (5.6 ng at 8×10^3 c.p.m. ng^{-1}) and various concentrations of unlabelled vinculin. Incubation was in 150 mM NaCl, 0.1% NaN_3 , 0.1% β -mercaptoethanol, 50 mM Tris-HCl pH 7.5, containing 5 mg ml^{-1} bovine serum albumin. Rabbit anti-talin antiserum³ ($1 \mu\text{l}$) was added to each sample and the incubation continued on ice for a further 12 h. At this point, $20 \mu\text{l}$ of a 10% suspension of protein A-bearing staphylococci¹⁹ were added per assay and incubated for a further 12 h. The incubation was stopped by adding 2 ml of the incubation buffer, followed by a centrifugation at $3,000 \text{ g}$ for 10 min. The supernatant was discarded and the pellets resuspended in 2 ml of the same buffer. They were centrifuged as before and the amount of ^{125}I -vinculin in the pellets measured in a γ -counter. Controls were performed in which the talin was omitted and in these conditions less than 3% of the total counts bound, and these were subtracted from each measurement. Each point represents the mean of two experiments. Inset: The data were also plotted according to the Scatchard equation⁸: $B/F = N/K - B/K$, where $B = \text{mol vinculin bound per mol talin}$, F is the concentration of unbound vinculin (nM), N is the number of binding sites (mol mol^{-1}), and K is the equilibrium constant. From the extrapolated line (-----), N is ~ 0.28 molecule of vinculin per molecule of talin, and the dissociation constant, K_d is $\sim 10^{-8}$ M.

of talin with vinculin suggests a potential role in such a chain of attachment for these two proteins. Talin, however, does not bind to actin as judged by sedimentation experiments with F-actin (our unpublished results). Similarly, although various studies have indicated an interaction between vinculin and actin¹⁰⁻¹², it has recently been reported that the effect of vinculin preparations on F-actin viscosity can be separated from vinculin by chromatography on cation exchange columns¹³. We have confirmed that this activity in vinculin preparations is due to a contaminant and not caused by vinculin (B. M. Paschal and K.B., unpublished results). If an attachment role for these two proteins is to be established, it becomes important both to determine whether vinculin interacts with actin at all and also to look for other proteins, particularly in the plasma membrane, that interact with the talin-vinculin complex.

We thank Dr J. Otto for sending us a preprint of her work before publication; Dr M. Beckerle for critically reading the manuscript; and Ms L. Connell and Mr B. M. Paschal for help with many of the experiments. We thank the secretaries of the Department of Anatomy for preparation of the manuscript. This work was supported by grants from the NIH (GM 29860), the Muscular Dystrophy Association and the American Cancer Society.

Received 1 November 1983; accepted 8 February 1984.

- Geiger, B. *Cell* **18**, 193-205 (1979).
- Burridge, K. & Feramisco, J. R. *Cell* **19**, 587-595 (1980).
- Burridge, K. & Connell, L. *J. Cell Biol.* **97**, 359-367 (1983).
- Burridge, K. & Connell, L. *Cell Motility* **3**, 405-417 (1983).
- Otto, J. *J. Cell Biol.* **97**, 1283-1287 (1983).
- Feramisco, J. R., Smart, J. E., Burridge, K., Helfman, D. M. & Thomas, G. P. *J. biol. Chem.* **257**, 11024-11031 (1982).
- Siliciano, J. D. & Craig, S. W. *Nature* **300**, 533-535 (1982).
- Scatchard, G. *Ann. N.Y. Acad. Sci.* **51**, 660-672 (1949).

9. Branton, D., Cohen, C. M. & Tyler, J. *Cell* **24**, 24–32 (1981).
10. Jockusch, B. M. & Isenberg, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3005–3009 (1981).
11. Wilkins, J. A. & Lin, S. *Cell* **28**, 83–90 (1982).
12. Burrage, K. & Feramisco, J. R. *Cold Spring Harb. Symp. quant. Biol.* **46**, 587–597 (1982).
13. Evans, R. R. & Robson, R. M. *J. Cell Biol.* **97**, 282a (1983).
14. Feramisco, J. R. & Burrage, K. *J. biol. Chem.* **255**, 1194–1199 (1980).
15. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
16. Burrage, K. *Meth. Enzym.* **50**, 54–64 (1978).
17. Hunter, W. M. & Greenwood, F. C. *Nature* **194**, 495–496 (1962).
18. Burrage, K. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4457–4461 (1976).
19. Kessler, S. W. *J. Immun.* **115**, 1617–1624 (1975).

Isolation and characterization of human calcitonin gene-related peptide

Howard R. Morris, Maria Panico, Tony Etienne,
John Tippins, Samia I. Girgis* & Iain MacIntyre*

Department of Biochemistry, Imperial College,
London SW7 2AZ, UK

* Royal Postgraduate Medical School, Ducane Road,
London W12 0HS, UK

The rat calcitonin gene has recently been shown to encode a novel peptide (rat calcitonin gene-related peptide, rCGRP) thought to be produced in nervous tissue after tissue-specific RNA processing^{1,2}. This peptide has so far been identified only in rat tissue, by immunocytochemistry and immunoassay. We now report the isolation of a related (89% homology) peptide from human tissue (hCGRP) which we have sequenced using a novel mass spectrometric approach, fast atom bombardment (FAB) mapping^{3–5}. The human peptide differs significantly from the predicted rCGRP structure in four positions in the amino acid sequence (three effecting charge changes), and the presence of a disulphide bridge and an amide, surmised in the rat work, is proven in the hCGRP molecule. hCGRP was present in plasma from 10 patients with medullary thyroid carcinoma (MTC) and in 6 MTC tumours removed at surgery, suggesting the tissue distribution may differ from that in the rat where the peptide is reported to be absent from thyroid tissue⁶. hCGRP is shown to have biological activity and it is possible that its presence in MTC plasma may be responsible for some of the symptoms in this disease.

Recently it has been reported that the rat calcitonin gene encodes a highly active peptide that is produced by tissue-specific RNA processing^{1,2,6}. The rat peptide has been identified by immunocytochemistry within the central and peripheral nervous systems but is thought to be largely absent from the thyroid. There has previously been no evidence that a similar peptide is present in man.

To extract hCGRP, an antibody raised against Tyr⁰ rCGRP (28–37) (Peninsular Laboratories) coupled to ovalbumin was used to detect a cross-reacting peptide from 15 g of human MTC tissue, obtained at operation. The tissue was homogenized in acid⁷ and the supernatant adsorbed to Waters C-18 Sep-Pak cartridges which were eluted with methanol:1% trifluoroacetic acid (TFA) (4:1). The lyophilized extract was further purified on a column of SP-Sephadex C-25 (83 × 1.5 cm) using a linear gradient of 0–1.0 M NaCl in 30% acetic acid. hCGRP, calcitonin and katecalcin were detected by radioimmunoassay. hCGRP eluted as a single peak which was re-chromatographed by HPLC on a Spherisorb 5 μ ODS column (10 × 0.46 cm) using a gradient of 40–90% methanol in 0.2% TFA, overall recovery 45%.

The structure of human CGRP was determined by the FAB mapping strategy developed in this laboratory^{3–5,8,9}. A FAB map of the digest of 2 nmol hCGRP gave the spectrum shown in Fig. 1. Four tryptic peptides were observed, only two of which correspond in mass to peptides in the predicted rCGRP sequence at m/z 546 (Ser-Gly-Gly-Val-Val-Lys) and m/z 729 (Leu-Ala-Gly-Leu-Leu-Ser-Arg). In other experiments these pep-

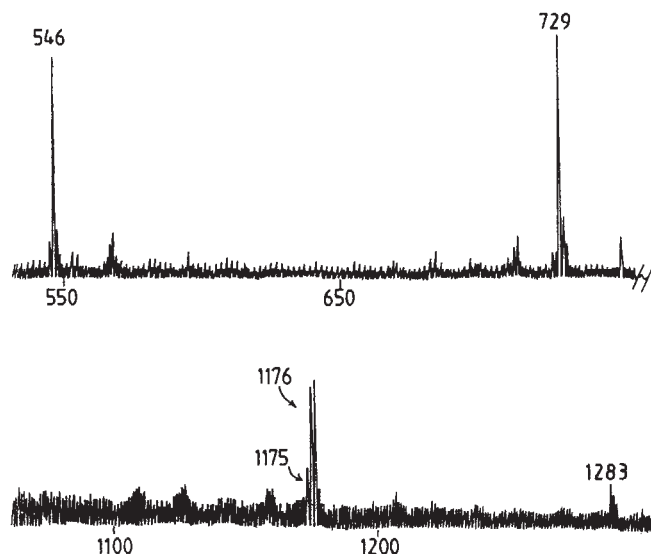


Fig. 1 FAB map of a tryptic digest of hCGRP in 5% acetic acid as solvent and in a matrix of glycerol and monothioglycerol. The spectra were recorded on a VG ZAB HF instrument using an M-Scan FAB gun operating at 20 μ A beam current at 10 kV.

tides were isolated and sequenced by FAB analysis, carboxypeptidase B and chymotryptic digestion and amino acid analysis (this allowed assignment of leucine rather than isoleucine which has the same mass difference of 113 AMU).

The two remaining unidentified peptides in Fig. 1 have quasimolecular ions at m/z 1,175 and 1,176, respectively. The signal at m/z 1,177 was deduced to be the di-thiol of an intra-S-S-bridged peptide at m/z 1,175 and the signal at m/z 1,283 was assigned to opening of the disulphide by the thioglycerol used in the FAB solvent matrix. Reduction and carboxymethylation of the tryptic digest followed by FAB mapping shifted the 1,175 signal to m/z 1,291 in the negative ion spectrum, confirming the presence of two cysteines in the molecule. That m/z 1,175 originated from the N-terminus of hCGRP was proven by first acetylating the intact molecule and then digesting with trypsin for a restricted reaction time of 10 min followed by FAB mapping. The m/z 1,175 signal incorporated an α -N-acetyl group and shifted to m/z 1,217, whereas the m/z 1,176 signal remained unchanged; the thioglycerol adduct ion also moved from m/z 1,283 to m/z 1,325, confirming the earlier interpretation. A further signal at m/z 1,927 also carrying the acetyl label showed that the order of the N-terminal tryptic peptides was the 1,175 peptide followed by the 729 peptide.

The mass of m/z 1,176 indicated a truncated sequence for hCGRP compared with a calculated C-terminal tryptic peptide mass for rCGRP of m/z 1,395. However, an experiment to determine the intact molecular weight of hCGRP gave a value for the $M+H^+$ signal of 3,786 (compared with 3,803 for rCGRP). The mass difference between this and the sum of the four tryptic peptides observed (3,569) showed that a small tryptic peptide remained unaccounted for and that a new tryptic cleavage point was present in hCGRP, close to the C-terminus.

Table 1 Amino acid ratios for HPLC peak tube fractions corresponding to peptides detected by analytical HVPE

HPLC fraction	Ratio of amino acids
Fraction 20	Asx _{0.48} Ser _{0.79} Gly _{1.79} Val _{1.58} Lys _{1.0}
Fraction 28	Ala _{1.08} Phe _{1.0}
Fraction 60	Asx _{1.02} Thr _{3.06} Ala _{1.64} Val _{1.02} His _{1.14} Arg _{1.0}
Fraction 72	Asx _{2.72} Thr _{0.97} Ser _{1.08} Gly _{1.3} Val _{1.97} Phe _{0.93} Lys _{1.0}
Fraction 76	Asx _{0.29} Ser _{1.22} Gly _{1.34} Ala _{1.04} Leu _{2.93} Arg _{1.0}

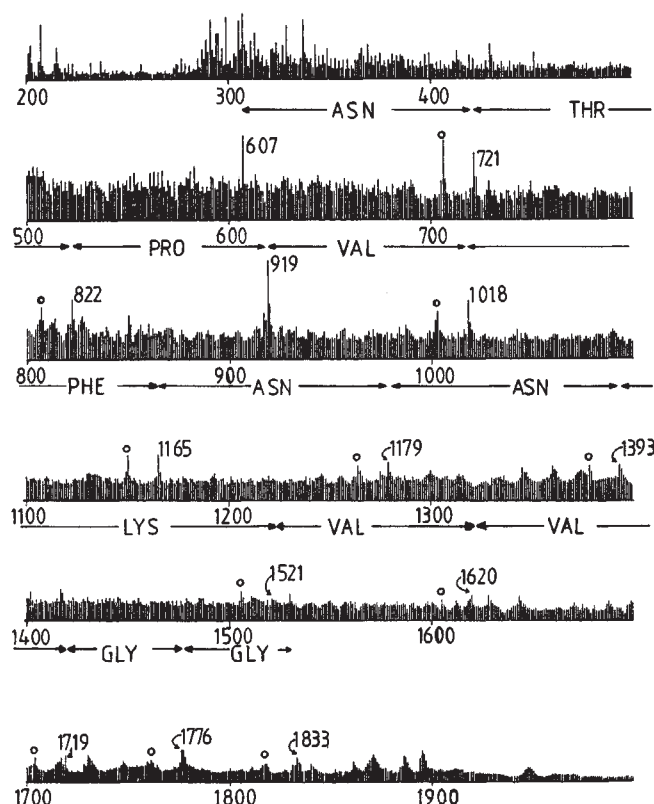
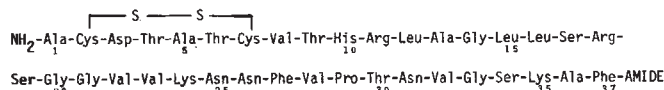


Fig. 2 Low mass region of the FAB mass spectrum of intact hCGRP recorded in the same conditions as Fig. 1. For clarity N-terminally derived signals (shifting by 42 AMU on acetylation) have not been numbered. Sequence ion (ammonium) signals are often accompanied by alkyl species 15 AMU lower (labelled o), of equal or higher intensity at the higher masses. Other signals of similar signal to noise ratio are either N-terminal sequence ions or do not have calculated mass differences corresponding to amino acids.

The low mass region of the FAB spectrum of intact hCGRP (Fig. 2) shows C-terminal sequence ions at m/z 607, 721, 822, 919, 1,018, 1,165, 1,279, 1,393, 1,521, 1,620, 1,719, 1,776 and 1,833. The mass differences define the C-terminal sequence of hCGRP as -Gly-Gly-Val-Val-Lys-Asn-Asn-Phe-Val-Pro-Thr-Asn-X where X is the C-terminus of the molecule and has a molecular weight of 605 (C-terminal sequence ions are observed 2 AMU higher in Fig. 2 due to protonation and rearrangement^{3,10}). These data also provide the necessary overlap between m/z 546 and the m/z 1,176 tryptic peptides in the FAB map (Fig. 1), and indicate (via the mass of 1,176) that a new tryptic cleavage site has been created by the presence of lysine at position 35 rather than the glutamic acid in rCGRP.

Six nanomoles of remaining hCGRP were then digested with trypsin and the peptides purified by HPLC using an acetic acid/*n*-propanol gradient elution¹¹. HPLC column fractions were screened analytically (~2 nM samples) by high voltage paper electrophoresis (HVPE) at pH 6.5, where five peptides were visualized by staining with Cd-ninhydrin. Amino acid analysis (Table 1) and dansylation of the purified peptides revealed that position 1 in the N-terminal (m/z 1,175) peptide is alanine rather than the serine predicted for rCGRP, and that a dipeptide Ala-Phe was released on tryptic digestion, confirming the assignment of Lys in position 35 of the sequence. Finally, the N- and C-terminal FAB fragmentation data of each of the HPLC-isolated tryptic peptides together with the FAB map (Fig. 1) and the spectrum of the intact hCGRP (Fig. 2) led to the assignment of the overall structure of hCGRP shown below:



The C-terminal amide was assigned definitively from the m/z 607 mass of a C-terminal fragment (Fig. 2; the mass would have been 608 if the molecule had possessed a C-terminal carboxyl group). This assignment is supported by the mobility of the Ala-Phe dipeptide on pH 6.5 HVPE, indicating a charge of +1. The presence of a disulphide bridge between residues 2 and 7, which can only be surmised from the predicted rat structure, is proven for hCGRP by the mass of both the N-terminal tryptic peptide (m/z 1,175, Fig. 1) and the intact molecule.

It had been anticipated from the rat studies that if hCGRP existed, it would not always be present in medullary thyroid carcinoma. However, we detected it in each of six cases of MTC removed at surgery and in plasma of 10 further cases (Table 2). It is not known whether the peptide is present in normal thyroid and circulates in normal subjects although preliminary studies in one normal thyroid obtained at autopsy suggest that CGRP is present. Our evidence implies that the distribution of hCGRP differs from that reported for the rat peptide although it should be noted that the rat studies were based on immunological evidence only and that our study rests on rigorous chemical characterization of the peptide from MTC.

Preliminary data on biological experiments, which were limited by lack of peptide, show hCGRP can increase the rate and force of contraction of isolated rat auricles, thus extending experiments with the rat peptide⁶.

There are interesting theoretical and practical implications of this work. First, the human calcitonin gene may well prove a convenient model system for the study of tissue-specific gene expression. Second, immunoassay of the peptide may be of diagnostic use in MTC and in cases with high levels of peptide it should be possible to determine whether some symptoms can be attributed to it. In any event this novel human peptide may prove to have an unexpected role in normal human physiology.

We thank Gamini Aleyasekera, Clair Myers and Tim Arnett for assistance with CGRP coupling, raising of antibodies and chromatography, and the MRC for financial support.

Table 2 Immunoreactive peptide levels in tumour tissue (a) and plasma (b) from patients with medullary thyroid carcinoma

a Patient	Calcitonin	Katacalcitonin	Calcitonin gene-related peptide
M.W.	470	9.0	8.4
R.B.	279	26	0.26
T.K.	323	19.6	0.21
D.	30	10.2	1.37
L.Y.	13	25	5.1
C.	12	6.5	0.35
b Patient	Calcitonin	Calcitonin gene-related peptide	
P.B.	70	1.2	
G.	24	1.2	
R.B.	47	4.0	
R.	27	0.84	
D.D.	414	10.5	
P.M.	42	2.0	
R.C.	82	10.5	
H.	10	0.84	
W.G.	2.7	1.3	
O'N.	9.7	1.44	

a, Concentration of immunoreactive calcitonin, katacalcitonin and calcitonin gene-related peptide in MTC tissue extract expressed as nanomol per g wet weight. b, Plasma levels of immunoreactive calcitonin and calcitonin gene-related peptide in MTC patients expressed as nanomol per l.

Received 9 November 1983; accepted 7 February 1984.

1. Amara, S. G., Jonas, V., Rosenfeld, M. G., Ong, E. S. & Evans, R. M. *Nature* **298**, 240–244 (1982).
2. Rosenfeld, M. G. *et al.* *Nature* **304**, 129–135 (1983).
3. Morris, H. R. *et al.* *Biochem. biophys. Res. Commun.* **101**, 623–631 (1981).
4. Morris, H. R., Panico, M., Karplus, A., Lloyd, P. E. & Riniker, B. *Nature* **304**, 643–645 (1982).
5. Morris, H. R., Panico, M. & Taylor, G. W. *Biochem. biophys. Res. Commun.* **117**, 299–305 (1983).
6. Fisher, L. A. *et al.* *Nature* **305**, 534–536 (1983).
7. Bennett, H. P. J., Hudson, A. M., Kelly, L., MacMartin, C. & Purdon, G. E. *Biochem. J.* **175**, 1139–1141 (1978).
8. Morris, H. R., Dell, A. & McDowell, R. A. *Biomed. Mass Spectrom.* **8**, 463–473 (1981).
9. Barber, M., Bordoli, R. S., Sedgwick, R. D. & Tyler, A. N. *Chem. Commun.*, 325–327 (1981).
10. Williams, D. H., Bradley, C., Bojesan, G., Santikarn, S. & Taylor, L. C. E. *J. Am. chem. Soc.* **103**, 5700–5704 (1981).
11. Morris, H. R., Etienne, A. T., Dell, A. & Albuquerque, R. J. *Neurochem.* **34**, 574–582 (1980).

Transmission of the polyoma virus middle T gene as the oncogene of a murine retrovirus

Daniel J. Donoghue*, Crystal Anderson*,
Tony Hunter† & Paul L. Kaplan†

* Department of Chemistry, University of California, San Diego,
La Jolla, California 92093, USA

† Molecular Biology and Virology Laboratory, The Salk Institute,
PO Box 85800, San Diego, California 92138, USA

Polyoma virus is a papovavirus that productively infects mouse cells. In cells of other species, such as rat cells, polyoma virus is virtually unable to replicate, and a small proportion of infected cells become stably transformed. The ability of polyoma virus to transform infected cells is determined by genes that encode the large, middle and small T antigens^{1–3} and which are found in the early region of the virus genome. We have inserted the transforming region of polyoma virus into a murine leukaemia virus (MLV) vector, to generate a replication-defective transforming retrovirus which for the first time allows efficient transformation of mouse cells by the polyoma virus middle T gene. During the life cycle of this recombinant virus the intervening sequence present in the original polyoma virus middle T gene was removed. The recombinant virus that we have constructed is analogous to other acutely transforming retroviruses, and demonstrates that the polyoma middle T gene is a dominant transforming oncogene.

Polyoma virus middle T antigen is found associated with the plasma membrane of virus-infected cells^{4–6}, and in immune complexes apparently undergoes autophosphorylation on tyrosine residues^{7,8}. However, middle T may not possess intrinsic kinase activity—recent data suggest that its phosphorylation may be due to an association with the host cell protein p60^{c-src}⁹. A variety of experiments^{10–12} demonstrate that middle T is responsible for the oncogenicity of polyoma virus, although ancillary roles have been suggested¹³ for small T and for the N-terminal portion of large T.

The recombinant clone pDD101 was constructed as shown in Fig. 1. A clone of an integrated MLV provirus¹⁴ was altered so that it could serve as a vector for *Hind*III restriction fragments. A 3.0 kilobase (kb) *Hind*III fragment of polyoma DNA was inserted into the MLV vector, orientated so that polyoma virus early transcription would be in the same direction as MLV transcription. This inserted fragment contains the early promoter, the entire coding region for middle T and small T, a portion of the large T coding region and all donor and acceptor splice sites for T antigen mRNAs. As a consequence of the substitution of polyoma viral sequences in place of MLV *pol* and *env* gene information, pDD101 could only give rise to a replication-defective retrovirus.

NIH3T3 cells were transfected with DNA of clone pDD101 using the calcium phosphate-DNA co-precipitation technique¹⁵,

and scored for foci of transformed cells. As shown in Table 1, pDD101 was biologically active for focus formation. pDD101 was also co-transfected with a clone of MLV, designated p836. These transfected cells were expected to make infectious MLV and thereby allow packaging of virion RNA of the replication-defective virus encoded by pDD101. Culture fluids were collected from the transfected cells of each sample in Table 1 and assayed for focus-forming activity. Only the cells co-transfected with both pDD101 and p836 yielded transmissible focus-forming activity. These results indicate that in the presence of helper MLV, cells transfected with pDD101 yielded a transmissible virus, designated PyMLV, which induced *de novo* focus formation.

The inserted fragment of polyoma DNA in pDD101 contains the splice donor and acceptor sequences for the middle T gene. We wished to determine whether the middle T intron was removed from PyMLV during propagation as a retrovirus. Primer extension sequencing was employed to determine directly the sequence across the middle T splicing sites of PyMLV, using the protocol outlined in Fig. 2. A sequencing gel of the primer extension reaction products is shown in Fig. 3a. This sequence analysis shows that the middle T intron was removed faithfully during propagation of PyMLV and selection for focus formation.

In order to express a functional middle T protein from the early region of the polyoma virus genome, an intron of 62 bases must be removed from the early region transcript. Although the polyoma DNA cloned into pDD101 contained this intron, the resulting PyMLV does not. Thus, the polyoma middle T sequences in PyMLV can be viewed as a cDNA copy of its spliced RNA. In this regard, PyMLV is analogous to other transforming retroviruses examined to date, which lack introns in their viral oncogenes. It was previously shown¹⁶ that an intron can be specifically removed from an α -globin gene inserted into a spleen necrosis virus vector. The present study extends this finding by showing that selection for one of several splicing events can occur, as indicated by the fact that PyMLV (encoding only middle T) was only one of several viruses which might have arisen from pDD101 by splicing events.

In order to demonstrate that PyMLV directs the synthesis of authentic middle T antigen, cell lines infected with either PyMLV or polyoma virus were examined for the presence of middle T. Cells were labelled with ³⁵S-methionine, and cell lysates were immunoprecipitated with polyoma anti-tumour serum. Immunoprecipitates were analysed by SDS-polyacrylamide gel electrophoresis and detected by fluorography¹⁷. Figure 3b, lane 4, shows the immunoprecipitate of a lysate

Table 1 Transformation of NIH3T3 cells

Transfected DNA	Foci per pmol DNA	Focus-forming units per ml filtered supernatant from transfected cells
Mock	0	< 5.0 × 10 ²
pPyMT1	5.2 × 10 ³	< 5.0 × 10 ²
pDD101	7.0 × 10 ³	< 5.0 × 10 ²
pDD101 + p836	3.7 × 10 ⁴	1.0 × 10 ⁵

Plasmid DNAs were transfected into NIH3T3 cells using the calcium phosphate-DNA coprecipitation technique¹⁵. Duplicate 1 µg samples of each DNA were assayed in the presence of 15 µg of calf thymus carrier DNA. Each sample was applied as a calcium phosphate-DNA co-precipitate to a 5 cm dish seeded about 24 h previously with 2.5 × 10⁵ NIH3T3 cells. Each dish was subsequently trypsinized, and the transfected cells distributed to four 10 cm dishes. The focus assay was scored 7 d after transfection. Four days after transfection, supernatants were collected from the transfected plates and filtered through a Nalgene 0.45 µm filter. Serial dilutions of the culture fluids were assayed for transforming virus on plates of NIH3T3 cells. Virus adsorption to subconfluent monolayers was for 1 h, in the presence of 8 µg ml⁻¹ of polybrene.

Fig. 1 Construction of pDD101. p836 consists of a large *Eco*RI fragment, containing a biologically active MLV provirus¹⁴, cloned onto the *Eco*RI site of pBR322. It was necessary to remove the *Hind*III and *Cla*I sites of the pBR322 portion of p836, which was done as follows: DNA of pBR322 was digested with *Hind*III and *Cla*I, after which the short protruding single-stranded ends were filled in using the Klenow fragment of *Escherichia coli* DNA polymerase I. The DNA was ligated and transfected into competent *E. coli* and the plasmid pDD61 was isolated, which lacks the *Cla*I and *Hind*III restriction sites. DNAs of p836 and pDD61 were cleaved with *Eco*RI, ligated together and transfected into competent bacteria. The plasmid pDD68 was isolated, which contains the large *Eco*RI fragment of p836 cloned into pDD61. pDD68 actually has two *Cla*I restriction sites, although one of these (at MLV position 4980 in the sequence of ref. 18) is methylated in DNA grown in *dam*⁺ bacteria, in which case it is not cleaved by *Cla*I. DNA of pDD68 was digested with *Cla*I and the single-stranded ends were filled in using the Klenow fragment of *E. coli* DNA polymerase I. This DNA was ligated with the *Hind*III linker CCAAGCTTGG, which was previously phosphorylated using T₄ polynucleotide kinase. After transfection, the plasmid pDD85 was identified in which the *Cla*I site was converted to a *Hind*III site. To construct pDD101, equal amounts of pDD85 DNA and polyoma form I DNA were mixed, digested with *Hind*III, and ligated. After transfection, pDD101 was isolated, which contains the large *Hind*III fragment of pDD85 and the 3.0 kb *Hind*III fragment of polyoma virus. The orientation of the polyoma fragment in pDD101 was confirmed to be as shown in the figure by digestion with *Eco*RI, which cleaves asymmetrically within the polyoma *Hind*III fragment.

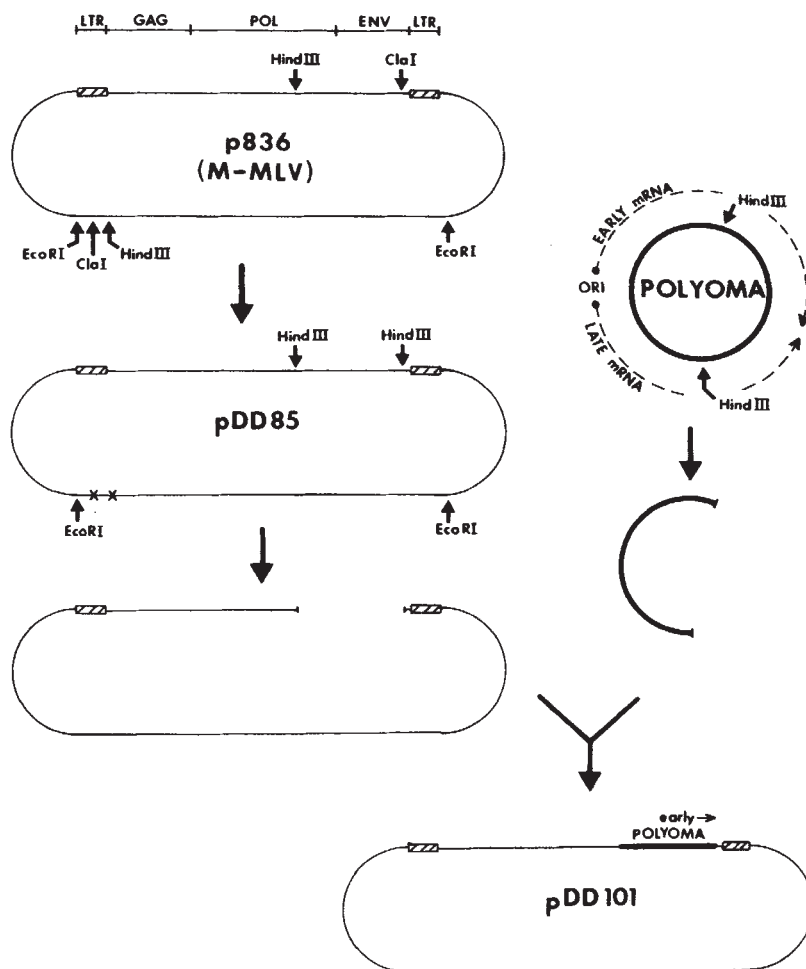
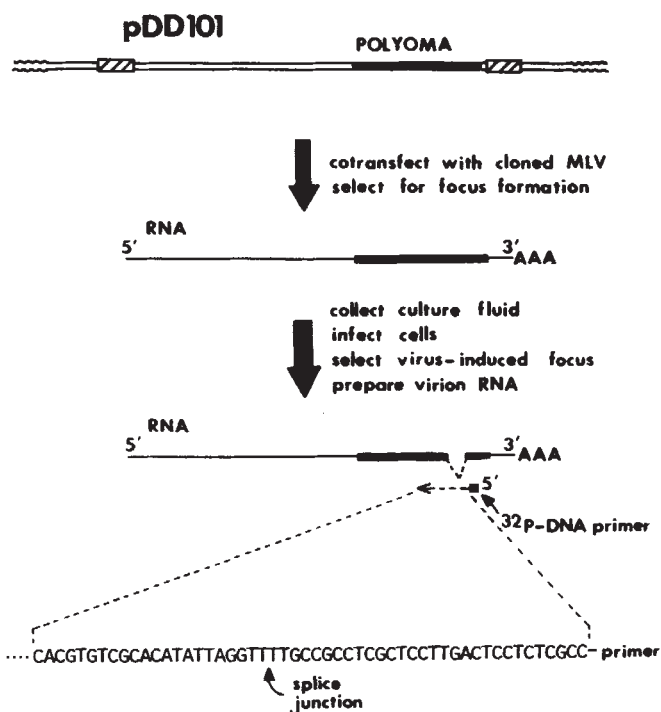


Fig. 2 PyMLV isolation and primer extension nucleotide sequencing. NIH3T3 cells were co-transfected with pDD101 and p836. The filtered culture fluid of these plates, which exhibited many foci, was subsequently assayed on uninfected NIH3T3 cells. Cell lines were prepared from several foci produced by infection at limiting multiplicity of infection, and were designated PyMLV-1, PyMLV-2 and so on. Typical virus titres for these cell lines were about 1×10^5 f.f.u. ml⁻¹ assayed on NIH3T3 cells, and about 1×10^5 XC p.f.u. ml⁻¹ assayed on XC cells¹⁹. Viruses were prepared from roller bottles seeded with PyMLV-1 cells by rapid harvesting (every 8 h) over a period of about one week. Viruses were concentrated and purified by sucrose gradient centrifugation according to standard procedures²⁰. The primer for primer extension sequencing consisted of a 50 base pair *Hae*III-*Rsa*I fragment of the polyoma virus genome which lies just downstream of the middle T splice acceptor site (nucleotides 852-902, sequence of Deininger *et al.*²¹). This fragment was first subcloned as a unique *Eco*RI-*Xba*I restriction fragment by restriction site reconstruction²². The primer fragment was 5'-³²P-labelled using polynucleotide kinase²³ and the strands were separated by electrophoresis after denaturation in a 6% polyacrylamide sequencing gel. The slower-migrating strand was determined empirically to hybridize to PyMLV RNA, which was prepared from sucrose-gradient purified virions by phenol extraction. Hybridization of the primer with the RNA was in a buffer of 62% formamide, 0.4 M NaCl, 20 mM PIPES pH 6.4, 2 mM EDTA. The nucleic acids were first placed at 85 °C for 5 min to denature the RNA, and then placed in a 65 °C waterbath, which was allowed to slowly cool to room temperature overnight. The RNA-DNA hybrids were then precipitated from the hybridization mixture with ethanol, and redissolved in the buffer described by Bina-Stein *et al.*²⁴. This was split into four aliquots, each of which was supplemented with one of the four dideoxynucleoside triphosphates (ddNTPs). The ddNTP:dNTP ratios were those used previously²². AMV reverse transcriptase was added for 1 h at 40 °C. Reaction products were precipitated with ethanol, and analyzed by electrophoresis in a 6% polyacrylamide sequencing gel.



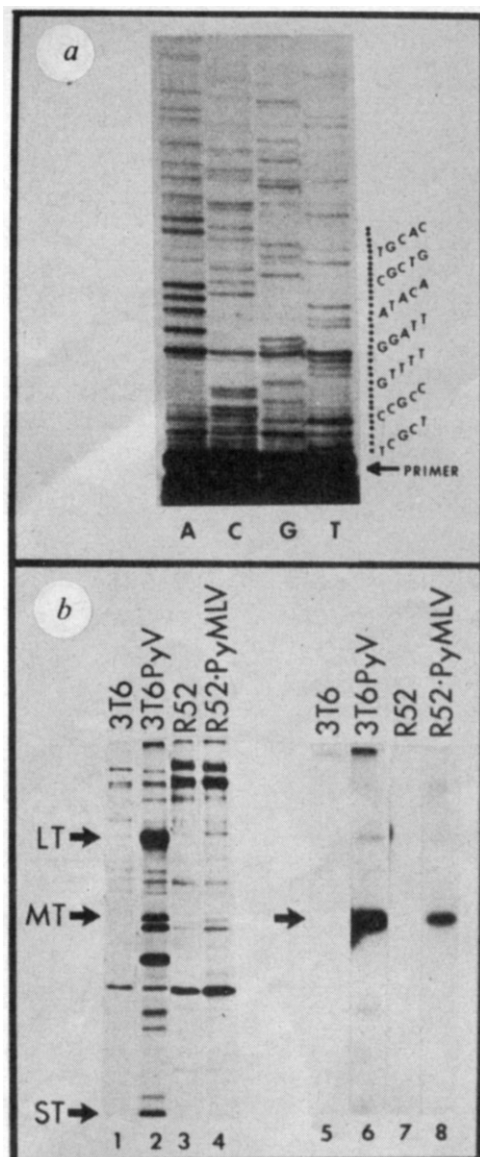


Fig. 3 Characterization of the PyMLV genome. *a*, The reaction products of primer extension nucleotide sequencing are shown after electrophoresis in a 6% polyacrylamide sequencing gel. The protocol is described in the legend to Fig. 2. The sequence shown here verifies the removal of the middle T intron from PyMLV virion RNA when compared to the expected sequence, which is shown in Fig. 2. There is some ambiguity in the sequence in the first several residues after the primer. This is due to a secondary structure feature of the primer fragment, which gives rise to a single-stranded DNA molecule with a hairpin formed by hybridization to itself, thereby creating a substrate for reverse transcriptase. This spurious hybridization gives rise to a 'shadow sequence' in the bottom of the gel, which terminates when the reverse transcriptase reaches the end of the hairpin formed by the single-stranded DNA fragment. Also note the 'strong stop' in the sequence at the end of the TTTT. This is characteristic of strings of T residues when sequenced by this method. *b*, SDS-polyacrylamide gel electrophoresis of proteins immunoprecipitated using polyoma anti-tumour serum. In lanes 1–4 *in vivo* ^{35}S -methionine-labelled proteins were prepared for immunoprecipitation as described by Eckhart *et al.*⁷, and detected by fluorography¹⁷. In lanes 5–8, aliquots of the same lysates were examined for *in vitro* kinase activity as described by Walter *et al.*⁸, and ^{32}P -labelled proteins were detected by autoradiography. Note the band of ^{35}S -methionine-labelled middle T in PyMLV(MLV)-infected REF52 cells (lane 4), which is also ^{32}P -labelled in the *in vitro* kinase assay (lane 8). REF52 is abbreviated R52. REF52 cells²⁵ were obtained from William Topp. A cloned focus of REF52 cells infected with PyMLV(MLV) was supplied by Brad Ozanne. Polyoma anti-tumour serum was prepared as the ascites fluid from Brown Norwegian rats bearing polyoma-induced tumours.

of a rat cell line REF52 infected with PyMLV. A band of ^{35}S -methionine-labelled middle-T is evident, which co-migrates with the band of middle T in lysates from polyoma virus-infected 3T6 cells (Fig. 3*b*, lane 2).

We also wished to examine whether middle T encoded by PyMLV was phosphorylated in an *in vitro* kinase assay, as is the case for authentic middle T⁷. Cell lysates were precipitated with polyoma anti-tumour serum, and then incubated with [γ - ^{32}P] ATP in a protein kinase assay. ^{32}P -labelled proteins were analysed by SDS-polyacrylamide gel electrophoresis and detected by autoradiography (Fig. 3*b*, lanes 5–8). A strong band of ^{32}P -labelled middle T was evident in the lysate of 3T6 cells infected with polyoma virus (lane 6), and also in the lysate of REF52 cells infected with PyMLV (lane 8). The identity of the presumptive ^{32}P -labelled middle T protein was confirmed by partial proteolysis mapping (data not shown).

These results demonstrate that PyMLV directs the synthesis of authentic polyoma middle T. When examined by ^{35}S -methionine labelling the amount of middle T present in PyMLV-infected cells, although comparable to that in transformed cells, was not as high as in polyoma virus-infected cells (data not shown). This may be attributable to the lower copy number of viral genomes in PyMLV-infected cells, as compared with cells acutely infected with polyoma virus. When examined by an *in vitro* kinase assay, the middle T present in PyMLV infected cells was significantly phosphorylated, as is authentic middle T antigen.

Normally, polyoma virus transforms cells at only a fraction of its theoretical efficiency, and only 1% or fewer of semi-permissive cells become permanently transformed. It has not been possible to rule out that the observed foci were derived from a small sub-population of susceptible cells. Infection of cells with PyMLV results in the transformation of essentially all the cells in the culture (data not shown). This demonstrates that the polyoma middle T gene, as is the case for other oncogenes transduced by retroviruses, can function as a dominant selectable oncogene.

We thank C. Tabin and R. A. Weinberg for p836 DNA, and D. Templeton and W. Eckhart for helpful discussions. P.L.K. acknowledges a postdoctoral fellowship from the Damon Runyon-Walter Winchell Cancer Fund. C.A. was supported by a Cell and Molecular Biology Training Grant (GMO7313). Portions of this work were funded by USPHS grants CA34456 and CA28458.

Received 20 September 1983; accepted 22 February 1984.

- Friedmann, T., Esty, A., Laporte, P. & Deininger, P. *Cell* **17**, 715–724 (1979).
- Soeda, E., Arrand, J. R., Smolar, N., Walsh, J. E. & Griffin, B. E. *Nature* **283**, 445–453 (1980).
- Treisman, R., Cowie, E., Favaloro, J., Jat, P. & Kamen, R. *Molec. appl. Genet.* **1**, 83–92 (1981).
- Ito, Y., Brocklehurst, J. R. & Dilbecco, R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4666–4670 (1977).
- Smart, J. E. & Ito, Y. *Cell* **15**, 1426–1437 (1978).
- Hutchinson, M. A., Hunter, T. & Eckhart, W. *Cell* **15**, 65–77 (1978).
- Eckhart, W., Hutchinson, M. A. & Hunter, T. *Cell* **18**, 925–933 (1979).
- Walter, G., Hutchinson, M. A., Hunter, T. & Eckhart, W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4025–4029 (1982).
- Courtneidge, S. A. & Smith, A. E. *Nature* **303**, 534–539 (1983).
- Treisman, R., Novak, U., Favaloro, J. & Kamen, R. *Nature* **292**, 595–600 (1981).
- Templeton, D. & Eckhart, W. *J. Virol.* **41**, 1014–1024 (1982).
- Carmichael, G. G., Schaffhausen, B. S., Dorsky, D. L., Oliver, D. B. & Benjamin, T. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3579–3583 (1982).
- Rassoulzadegan, M. *et al. Nature* **300**, 713–718 (1982).
- Hoffman, J. W. *et al. J. Virol.* **44**, 144–157 (1982).
- Graham, F. L. & van der Eb, A. J. *Virology* **52**, 456–467 (1973).
- Shimotohno, K. & Temin, H. M. *Nature* **299**, 265–268 (1982).
- Laskey, R. A. & Mills, A. D. *Eur. J. Biochem.* **56**, 335–341 (1975).
- Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. *Nature* **293**, 543–548 (1981).
- Rowe, W. P., Pugh, W. E. & Hartley, J. W. *Virology* **42**, 1136–1139 (1970).
- Papkoff, J., Hunter, T. & Beemon, K. *Virology* **101**, 91–103 (1980).
- Deininger, P. L., Esty, A., LaPorte, P., Hsu, H. & Friedmann, T. *Nucleic Acids Res.* **8**, 855–860 (1980).
- Donoghue, D. J. & Hunter, T. *Nucleic Acids Res.* **10**, 2549–2564 (1982).
- Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
- Bina-Stein, M., Thoren, M., Salzman, N. & Thompson, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 731–735 (1979).
- McClure, D. B., Hightower, M. J. & Topp, W. C. in *Growth of Cells in Hormonally Defined Media* (eds Sato, H. T., Pardee, A. B. & Sirbasku, D. 345–364 (Cold Spring Harbor, New York, 1982).

Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus

John Armstrong, Heiner Niemann*, Sjef Smeekens†‡, Peter Rottier† & Graham Warren

European Molecular Biology Laboratory, Postfach 10.2209, 6900 Heidelberg, FRG

* Institut für Virologie, Fachbereich Humanmedizin der Justus-Liebig Universität, Giessen, FRG

† Institute of Virology, Veterinary Faculty, State University of Utrecht, 3509 TD Utrecht, The Netherlands

In the eukaryotic cell, both secreted and plasma membrane proteins are synthesized at the endoplasmic reticulum, then transported, via the Golgi complex, to the cell surface¹⁻⁴. Each of the compartments of this transport pathway carries out particular metabolic functions⁵⁻⁸, and therefore presumably contains a distinct complement of membrane proteins. Thus, mechanisms must exist for localizing such proteins to their respective destinations. However, a major obstacle to the study of such mechanisms is that the isolation and detailed analysis of such internal membrane proteins pose formidable technical problems. We have therefore used the E1 glycoprotein from coronavirus MHV-A59 as a viral model for this class of protein. Here we present the primary structure of the protein, determined by analysis of cDNA clones prepared from viral mRNA. In combination with a previous study of its assembly into the endoplasmic reticulum membrane⁹, the sequence reveals several unusual features of the protein which may be related to its intracellular localization.

The coronaviruses are a diverse class of enveloped RNA viruses of considerable medical and agricultural significance; they also provide a model for the study of persistent viral infections (see ref. 10 for review). In contrast to many enveloped viruses, the coronavirus mouse hepatitis virus (MHV) A59 buds inside the cell, into the lumen of the endoplasmic reticulum¹¹⁻¹⁴. The assembled virion then appears to travel, via the Golgi complex, to the cell surface. Of the two viral membrane proteins, the smaller one, E1, is necessary for formation of the envelope, and is restricted to internal cell membranes; apparently it only reaches the cell surface as part of the budded virion^{12,13}. Thus, the E1 glycoprotein is potentially a convenient model for studying those features of a membrane protein that determine its arrest at a particular destination on the membrane transport pathway.

The mRNAs of MHV-A59 form a 'nested set': the seven RNAs share the 3' region of the positive-stranded genome, but extend to different lengths towards the 5' end¹⁵⁻¹⁸. From each RNA, only the 5' gene is translated^{19,20}. In addition, a non-coding 'leader' sequence of approximately 70 bases, from the 5' end of the genome, is common to the mRNAs^{18,21,22}. The E1 gene is second from the 3' end and is therefore translated from the second smallest mRNA, RNA 6 (refs 19, 20). The sequence of the 3'-terminal gene, encoding the viral nucleocapsid protein, has been determined previously^{23,24}.

Copy DNA clones spanning the E1 gene were prepared by two methods²³⁻²⁵ and sequenced in the vectors M13mp8 (ref. 26) or pEMBL8 (ref. 27) by the chain-termination method²⁸ (data available on request). A sequence of 780 nucleotides (Fig. 1), containing a single long open reading frame, precedes the coding region for the viral nucleocapsid protein. A leader of 76 nucleotides, almost identical to the leader of the smallest mRNA, RNA 7 (ref. 22), lies in front of the first potential initiator codon. Thus, the sequence in Fig. 1 represents the 5' end of RNA 6, encoding the E1 protein and starting at or near the extreme 5'-terminal nucleotide.

```
CCTATAAGAGTGATTGGCGTCGCTACGTACCCCTCAACTCTAAACTCTGTAGTTTAA
10          30          50
MetSerSerThrThrGlnAlaProGluProValTrpGlnTrpTh
ATCTAATCCAAACATTAGTAGTACTACTACGCCCCAGAGCCGCTCATCAATGGAC
70          90          110
rAlaAspGluAlaValGlnPheLeuLysGluTrpAsnPheSerLeuGlyIleIleLeuLe
GGCGACGAGGAGTCAATTCCTTAAGGAATGGAACCTTCGTTGGGCATTATACACT
130         150         170
uPheIleThrIleIleLeuGlnPheGlyTrpThrSerArgSerMetPheIleTyrValVa
CTTTATTACTATCATACTACAGTTCGGTTACAGAGCGTAGCATGTTTATTATGTTGT
190         210         230
lLysMetIleIleLeuTrpLeuMetTrpProLeuThrIleValLeuCysIlePheAsnCy
GAAATGCTAATCTTGTGGTTAATGTGGCTACTGATTGTTTGTGTTTCAATTG
250         270         290
sValTyrAlaLeuAsnAsnValTyrLeuGlyPheSerIleValPheThrIleValSerII
CGTGTATCGCTAAATAATGTGTATCTTGGATTTTCTATAGTGTATTACTAGTGCCAT
310         330         350
eValIleTrpIleMetTyrPheValAsnSerIleArgLeuPheIleArgThrGlySerTr
TGTAATCTGGATTATGTTTGTGTAATAGCATAAGGTTGTTTACAGGACTGGTAGCTG
370         390         410
pTrpSerPheAsnProGluThrAsnAsnLeuMetCysIleAspMetLysGlyThrValTy
GTGGAGCTTCAACCCGAAACAAACACCTTATGTGTATAGATAGAAAGTACCGTGTA
430         450         470
rValArgProIleIleGluAspTyrHisThrLeuThrAlaThrIleIleArgGlyHisLe
TGTTAGACCACTATTGAGGATTACCATACACTAACAGCCACTATTTCGTGGCCACCT
490         510         530
uTyrMetGlnGlyValLysLeuGlyThrGlyPheSerLeuSerAspLeuProAlaTyrVa
CTACATGCAAGGTGTTAAGCTAGGACCGGTTTCTCTTGTCTGACTGGCCGCTTATGT
550         570         590
lThrValAlaLysValSerHisLeuCysThrTyrLysArgAlaPheLeuAspLysValAs
TACAGTTCTCAAGGTGTCACACCTTTGCACCTATAAGCGCGCATTTAGACAAAGGTAGA
610         630         650
pGlyValSerGlyPheAlaValTyrValLysSerLysValGlyAsnTyrArgLeuProSe
CGGTGTAGCGGTTTGTCTGTTTATGTGTAAGTCCAGGTCGGAATTAACGACTGCCCTC
670         690         710
rAsnLysProSerGlyAlaAspThrAlaLeuLeuArgIle*
AAACAACCGAGTGGCGGACACCCGCTTGTGAGAACTAATCTAAACTTTAAGGATG
730         750         770
```

Fig. 1 Sequence of the E1 cDNA and protein extending to the initiator codon of the adjacent nucleocapsid gene²³. Proposed membrane-spanning regions are overlined.

Two versions were found, in two different clones, for the sequence immediately upstream from the E1 initiator codon. The shorter one is shown in Fig. 1; in the second clone, an additional copy of the pentanucleotide ATCTA was found between nucleotides 65 and 66, making the sequence similar to that of the region adjacent to the nucleocapsid gene of another strain of MHV²⁹. This difference could represent a mutation; alternatively, it may reflect heterogeneity in the normal mRNA population. Indirect support for the latter possibility comes from the observation that a RNase-T₁ oligonucleotide from this region of RNA 6, corresponding to the shorter sequence, was recovered in markedly lower yield than those from the rest of the molecule³⁰. This site represents the point of fusion between the 5' leader sequence and the coding portion of the RNA. The fusion is thought to occur by 'jumping' of the viral RNA polymerase to particular sites on its genome-length, negative-stranded template; the resumption of transcription then produces each of the subgenomic mRNAs^{22,31,32}. Thus, it seems possible that the polymerase may jump to more than one point on the template for each mRNA, generating variable numbers of the repeated pentanucleotide AUCUA in the resulting transcript.

Figure 1 shows the amino acid sequence encoded by the E1 gene. The predicted molecular weight of the protein is 26,000, slightly higher than that observed by gel electrophoresis^{19,33} but consistent with the unusual electrophoretic behaviour of this³³, and other, hydrophobic proteins. Several features of the protein, when assembled into membranes in the virus³³, or *in vitro*⁹, are reflected in the sequence. First, in contrast to the majority of membrane proteins, E1 is known to lack a cleaved 'signal peptide'⁹: the N-terminal region of the sequence contains no good candidate for a cleavage site³⁴. Second, the N-terminal region bears O-linked sugars^{35,36}, which, uniquely among viral proteins so far studied, are the only known post-translational modification to E1. Assuming that the terminal Met is removed³⁷, the N-terminal sequence is Ser-Ser-Thr-Thr, which is identical to the O-glycosylated amino terminus of M-type glycophorin A (ref. 38). The O-linked sugars of E1 are them-

‡ Present address: Molecular Cell Biology Group, State University of Utrecht, 3509 CH Utrecht, The Netherlands.

Seal echolocation?

THE evidence reported by Renouf and Davis¹ for echolocation by a 7-yr-old male harbour seal (*Phoca vitulina*) is not convincing enough to require a reassessment of earlier studies²⁻⁵ which found no experimental evidence for echolocation in pinnipeds.

The ability of the seal to swim across the 7.5 m diameter tank and locate a ring in the dark does not in itself indicate an echolocation ability. In field tests one of us (N. Sonafrank, R. Elsner and D.W., in preparation) has demonstrated the ability of a blindfolded spotted seal (*Phoca vitulina largha*) to find and surface through a 50-cm diameter hole in the ice from a starting hole over 30 m away. The seal produced no vocalizations during this sub-ice swimming and was equally successful in the presence of high levels of white background noise. Moreover, Oliver⁴ found a negative correlation between sound emission and improved navigational ability by a 3-yr-old male grey seal (*Halichoerus grypus*) swimming through different obstacle courses in the dark.

There are several problems with the experimental design of Renouf and Davis' ring discrimination experiment. First, even though the rings are visually identical, the dynamics of motion in the water of the weighted, air-filled ring will be quite different from that of the weighted, water-filled ring. Thus, the animal could discriminate between the rings visually by observing this motion in response to a pressure wave generated by the animal's swimming.

Since the majority of the experimental sessions (8-26) were conducted in the light and since the experimenter holding the targets by means of strings knew which one was the air-filled correct choice, the investigators have failed to rule out the possibility that they provided the animal with inadvertent cues. Finally, during the control sessions when both rings were water-filled, the learning curve, as indicated by the discrimination ratio in Fig. 2 of Renouf and Davis¹, is improving at a rate equal to or greater than that of the initial learning with the acoustically different rings during the experimental phase. Since the learning curve in the control situation had shown no sign of an asymptote, we cannot understand why the control phase was terminated after only 5 sessions compared with the 26 sessions of the previous experimental phase. We therefore find no justification for the conclusion that this was an adequate control test to confirm the previously demonstrated discrimination.

In a long-term study of vocalizations of spotted seals, we⁶ have recorded pulses similar to those reported by Renouf and Davis¹, although pulse trains were more common than single or double pulses. We find that a major change in frequency of occurrence of the pulse trains occurs on an annual cycle, peaking during the breeding seasons during the 6 yr we have

monitored these vocalizations. The field recordings by Renouf and Davis¹ were from a breeding colony, although they did not indicate the time of the year at which the recordings were made.

DOUGLAS WARTZOK

Purdue University,
Fort Wayne, Indiana 46805, USA

RONALD J. SCHUSTERMAN

California State University,
Hayward, California 94542, USA

JANET GAILEY-PHIPPS

Johns Hopkins University,
Baltimore, Maryland 21205, USA

1. Renouf, D. & Davis, M. B. *Nature* **300**, 635-637 (1982).
2. Evans, W. & Haugan, R. *Bull. Sth Calif. Acad. Sci.* **62**, 165-175 (1963).
3. Schusterman, R. J. in *Animal Sonar Systems: Biology and Bionics* (ed. Busnel, R.) 535-617 (Laboratoire de Physiologie Acoustique, Jouy-en-Josas, France, 1967).
4. Oliver, G. W. *Behaviour* **67**, 97-114 (1978).
5. Scronce, B. & Ridgway, S. in *Animal Sonar Systems* (eds Busnel, R. & Fish, J.) 991-993 (Plenum, London, 1980).
6. Wartzok, D., Gailey-Phipps, J., Schultz, N. & Beier, J. C. *Proc. 4th bien. Conf. Biol. of Marine Mammals*, 122 (University of California, Santa Cruz, 1981).

RENOUF REPLIES — I agree with Wartzok *et al.* that our study does not necessarily require the reassessment of those which failed to demonstrate echolocation. There is no reason to assume that what might be true for harbour seals would also hold for *Halichoerus*, *Zalophus* or *Pivipargha*, although the reverse logic should not be expounded either. My major concern was the fact that those interested in pinniped sensory function were tacitly 'accepting the null hypothesis' and concluding from negative evidence that seals do not use sonar.

I agree with most of the reservations expressed by Wartzok *et al.* about the study¹ and in fact alluded to many of them in the discussion. The discrimination ratios or the control tests were from 0.35 to 0.50 which Wartzok *et al.* feel is a significant trend. Apart from the fact that this difference did not prove to be statistically significant, a change in behaviour from mostly incorrect (0.35) to random (0.50) is not the same as the improvement shown by the seal during discrimination training from 0.63 to criterion. The problem of differential hydrodynamic motion of the rings occurred to me and was one of the reasons I switched to running in daylight. I wanted to make sure that the seal did not move both rings before making his decision. When we could see what was going on, we immediately released the first ring the animal touched as his choice. As far as we could tell, his swimming did not cause any ring motion (I assume Wartzok *et al.* meant displacement wave rather than pressure wave generated by the seal's movement)—which is not to say that what we could not detect that the seal could not. The other problem I still have is the relative paucity and quiet of the clicks the seal made during the discrimination. I postulated in the report¹ that perhaps this was his best solution to boundary reverberation. Since then I have replicated the study with the rings suspended in the

middle of the tank. The seal's discrimination ratio improved to 0.90; however, the clicks were still few and faint, but increased in peak energy to 21 kHz. The main difficulty is working in a relatively small tank where boundary reverberations are problematical. The vocalizations from the captive seals are weak compared with those I have obtained in the field, and so I am hoping to resolve this by building an ocean pen.

DEANE RENOUF

Department of Psychology,
Memorial University of Newfoundland,
St John's, Newfoundland,
Canada A1B 3X9

1. Renouf, D. & Davis, M. B. *Nature* **300**, 635-637 (1982).

DNA-binding proteins

In a recent News and Views article¹, Andrew Travers states that in a eukaryote '... a [specific] DNA-binding protein must have 10²-10³-fold higher affinity for its binding site than would be necessary in *Escherichia coli*... and that to achieve efficient occupancy the sequence specificity of the interaction relative to random binding must be 10³ times greater than that [of specific DNA-binding proteins found in *E. coli*]' Travers' conclusion is not true. In fact, based on what we know about two prokaryotic gene regulators, the *lac* and λ repressors, we expect that each would bind to its single specific operator equally efficiently whether that operator were in a bacterium or in a mammalian cell, provided that the concentration of the protein in the mammalian nucleus equalled that found in the bacterial nucleoid body. (The two prokaryotic repressors use somewhat different strategies to efficiently bind operator: λ repressor has lower affinities for operator and for random DNA than does *lac* repressor and, unlike the latter, it binds cooperatively to adjacent DNA sites. A bacterium typically contains about 10-20 *lac* repressors and the equivalent of about 100 dimers, the active species².)

Travers evidently believes that his conclusions follow from the facts that '... in a mammalian nucleus the concentration of a unique DNA-binding site [operator] is about 10² times per genome lower than in *E. coli* while the potential number of random DNA-binding sites is 10³-10⁴ times greater'. But the important factor is the concentration of repressor free to interact with operator, a quantity determined by the total repressor concentration and, in some cases, by the concentration of nonspecific DNA. Both λ and *lac* repressors, for example, have measurable affinities for nonspecific DNA and, especially in the latter case, this effectively decreases the free repressor concentration in a bacterium^{3,4}. The important point in the present context is that, according to Lin and Riggs, the total DNA concentration, and hence the concentration of nonspecific binding sites, is roughly the same in a mammalian cell nucleus and in a

bacterial nucleoid body³. Therefore, assuming equal concentrations of total repressor in a bacterium and in a mammalian nucleus, the fraction free to bind operator would be the same in each case. So long as the concentration of operator is low compared with the free repressor concentration, neither its absolute concentration nor the ratio of operator to nonspecific sites is relevant^{3,4}.

As might be expected from these considerations, the T-antigen of SV40, which binds to specific DNA sites with affinity roughly that of a prokaryotic regulator (a repressor), is present at about the same concentration in the nucleus of a transformed cell, that bears a single T-antigen gene, as is λ repressor in the nucleoid body of a λ lysogen (R. Tjian, personal communication). Evidently T-antigen finds its operator and efficiently regulates transcription in a transformed cell⁵.

Travers proposes that RNA polymerase II in eukaryotes is directed to promoter sites on DNA by protein(s) specifically bound to those sites. This idea is reasonable and consistent with emerging evidence^{6,7}; my point is that it does not follow from his arguments. Indeed, were Travers correct, the auxiliary binding protein, that which binds DNA and directs polymerase to its target, would itself be unable to occupy its site efficiently.

Perhaps it is worth recalling the example of *E. coli* RNA polymerase. This enzyme binds *in vivo* to many *E. coli* promoters and initiates transcription unaided by other DNA-binding proteins. However, the enzyme efficiently recognizes other *E. coli* promoters only if guided by DNA-bound positive regulatory proteins (see, for example, refs 8, 9). The manifest function of this activator-polymerase interaction is that it affords a means of gene control—certain genes are transcribed only if the positive regulator is active, whereas other genes are transcribed constitutively, by the same polymerase.

MARK PTASHNE

Department of Biochemistry
and Molecular Biology,
Harvard University, Cambridge,
Massachusetts 02138, USA

1. Travers, A. *Nature* **30**, 755 (1983).
2. Johnson, A. D. *et al. Nature* **294**, 5838 (1981).
3. Lin, S. & Riggs, A. D. *Cell* **4**, 107 (1975).
4. von Hippel, P. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **71**, 4808 (1974).
5. Tegtmeyer, P. *J. Virol.* **16**, 168 (1975).
6. Dynan, W. S. & Tjian, R. *Cell* (in the press).
7. Davison, B. L., Egly, J. M., Mulvihill, E. R. & Chambon, P. *Nature* **301**, 680 (1983).
8. Hochschild, A. *et al. Cell* **32**, 319 (1983).
9. Hawley, D. & McClure, W. R. *Cell* **32**, 327 (1983).

TRAVERS REPLIES—I recently suggested that eukaryotic RNA polymerases might locate a promoter by the initial recognition of a protein(s) prebound to this site rather than by the recognition of a specific DNA sequence¹. This suggestion was based on two arguments: first, that in eukaryotic cells the size of the genome severely limits possible mechanisms of promoter location; and second, that the

available experimental evidence suggests that DNA sequences common to most polymerase II promoters direct the binding of transcription factors other than RNA polymerase²⁻⁴. Ptashne suggests that the logic of my first argument is invalid.

In a seminal article Linn and Riggs⁵ discussed the limitations imposed by the size of the eukaryotic genome on the equilibrium binding of a regulatory protein to its DNA target site. They showed that if the number of molecules of *lac* repressor were the same in the prokaryotic cell and the eukaryotic nucleus, a condition which is also prerequisite for my own previous discussion and conclusions¹, the repressor would be unable to occupy its binding site efficiently. They suggested four possible mechanisms which would allow eukaryotes to overcome this problem of excess DNA. These mechanisms were an increased specificity of the protein-DNA interaction, a selective masking of competing DNA, tandemly repeated binding sites or, finally, an increased concentration of the DNA binding protein.

The apparent affinity of a DNA binding protein for its target can be described by the equation⁵

$$K_{\text{eff}} = K_{\text{RO}} + D_t K_{\text{RO}} / K_{\text{RD}}$$

where K_{RO} and K_{RD} are the equilibrium dissociation constants of the protein for specific and nonspecific sites, respectively, and D_t is the total concentration of nonspecific sites. The important parameter in the present discussion is the selectivity, $K_{\text{RD}}/K_{\text{RO}}$, which determines the partition of a binding protein between specific and nonspecific sites. For the particular case of *E. coli* RNA polymerase $K_{\text{RD}}/K_{\text{RO}}$ is $\sim 10^4$ for a strong promoter⁶, and $\sim 10^3$ for an average promoter. This means that in the context of a mammalian nucleus the effective number of random DNA sites would be equivalent to $\sim 6 \times 10^5$ strong polymerase binding sites. The number of polymerase II molecules in a mammalian nucleus is very approximately 5×10^4 (ref. 7). Clearly if polymerase II had the same value of $K_{\text{RD}}/K_{\text{RO}}$ as the bacterial enzyme and all the random DNA sites were competing with, say, 10^3 – 10^4 promoter sites promoter location would be kinetically inefficient and the occupancy at equilibrium would be low. To increase equilibrium occupancy one solution would be to increase $K_{\text{RD}}/K_{\text{RO}}$ as I argued previously. Alternatively a similar result could be achieved by masking at least 90% of the random binding sites. However the selectivity of the purified eukaryotic RNA polymerase is significantly less than that of the bacterial holoenzyme while a requirement for extensive masking would place constraints on the amount of DNA available for transcription at any one time. These considerations led to the suggestion¹ that promoter location would proceed more efficiently if the primary recognition of a promoter required a pro-

tein-protein interaction in place of a protein-DNA interaction.

Ptashne raises the relevant question as to how the protein which directs the polymerase to its target binds efficiently to its own DNA site. Kinetically, the problem faced by the auxiliary protein is entirely different from that for RNA polymerase. Whereas polymerase II can initiate *in vivo* at rates up to at least one every 4s the auxiliary protein-DNA complex could be extremely stable^{2,4}. Consequently the polymerase needs to locate a promoter rapidly by both two- and three-dimensional diffusion processes while the auxiliary factor(s) is not constrained to the same extent by kinetic limitations and once bound to its target need only remain.

Ptashne's own comments I believe to be based on a misconception and to be irrelevant to my original discussion. His argument is dependent on two crucial premises: first, that a DNA binding protein such as the *lac* repressor would bind to its operator equally efficiently in a mammalian nucleus and in a bacterium provided that the concentration of the protein in the mammalian nucleus equalled that found in the bacterial nucleoid; and second, that the total DNA concentration (D_t) in the nucleus is roughly the same as in the bacterial nucleoid body⁵. The equations derived by Linn and Riggs show that the concentration of *lac* repressor required for efficient (99.9%) repression in the mammalian nucleus, assuming no DNA masking, is 2×10^{-6} M. (The value of 4×10^{-7} M quoted by Linn and Riggs assumes that 80% of the DNA is unavailable to the repressor). From this number Ptashne's premises predict, assuming 10–20 repressor molecules per cell, that the volume of the *E. coli* nucleoid is ~ 0.008 – $0.016 \mu\text{m}^3$. The observed average volume of ~ 0.5 – $1 \mu\text{m}^3$ (refs 9, 10) is obviously substantially greater. Thus the DNA concentration in the nucleoid must be considerably lower than Ptashne's arguments assume. Perhaps significantly, at this lower DNA concentration the calculated repressor concentration required for efficient repression corresponds approximately to the value observed *in vivo*.

Finally I particularly thank Dr R. S. Jack for many constructive discussions during the development of the model for eukaryotic promoter location.

ANDREW TRAVERS

MRC Laboratory of
Molecular Biology,
Hills Road, Cambridge CB2 2QH, UK

1. Travers, A. *Nature* **303**, 755 (1983).
2. Davison, B. I. *et al. Nature* **301**, 680 (1983).
3. Dynan, W. S. & Tjian, R. *Cell* **32**, 669 (1983).
4. Fire, A., Samuels, M. & Sharp, P. A. *J. biol. Chem.* (in the press).
5. Lin, S.-Y. & Riggs, A. D. *Cell* **4**, 107 (1975).
6. Strauss, H. S., Burgess, R. R. & Record, T. M. *Biochemistry* **19**, 3504 (1980).
7. Schwarz, L. B. & Roeder, R. G. *J. biol. Chem.* **250**, 3221 (1975).
8. Le Meur *et al. Cell* **23**, 561 (1981).
9. Hecht, R. M., Taggart, R. T. & Pettijohn, D. E. *Nature* **253**, 60 (1975).
10. Van Ness, J. & Pettijohn, D. E. *J. molec. Biol.* **129**, 501 (1979).

In search of a subject

Anthony C.T. North

Biophysics. Edited by Walter Hoppe, Wolfgang Lohmann, Hubert Markl and Hubert Ziegler.

Springer-Verlag: 1983. Pp. 941. DM 168, \$65.90.

WHAT is biophysics? This question, asked by the editors of *Biophysics*, must also have been asked countless times of all us who profess to be biophysicists. To develop a sense of unity and to delineate boundaries is perhaps the main obstacle to be surmounted by anyone who seeks to write a core textbook on this discipline. Yet, the writing of such a text would in itself represent an essential stage in the definition of the subject. Biophysicists have long awaited such a statement — have the authors of *Biophysics* provided it?

The book is a compendium of articles to which no fewer than 68 authors have contributed under the supervision of four editors. This has enabled them to compile a survey of extraordinary breadth and frequently of overwhelming intellectual depth. The price that has had to be paid is unfortunately heavy and reveals itself in a lack of balance, of consistency and of coherence. The editors admit in their introduction the problems inherent in a multi-author work, but have not avoided all of them. Their principal failure is not to have established in the minds of the individual authors a clear idea of the intended readership, for they have given us a book parts of which are ideal material for undergraduate use, but other sections of which are likely to be of value to no more than a score of research groups throughout the world.

After a brief and useful introduction to the structure of cells, the book commences with structure at the molecular level and progresses generally to higher levels of organization and complexity. The chapter on the chemical structure of biologically important macromolecules is a generally satisfactory account of protein structure, apart from minor errors such as the statement that each subunit of haemoglobin binds one *atom* of oxygen. This chapter, however, fails to inform the reader of the controversies that have developed in recent years over the molecular structure of DNA, which have led to our current understanding that DNA is highly polymorphic, that it can adopt a number of regular conformations of well-defined geometries (not just the classical β -form described here) and that transitions between these forms are probably of the greatest importance in the control of gene expression. More remarkably, neither here nor elsewhere in the book is there any discussion of the chemical structures and properties of polysaccharides in general or of cellulose in particular, the most abundant of all biological polymers.

Biophysics is concerned both with gaining an understanding of the physical basis of the processes and structures essential to living beings and with the application of physical techniques to studying biological systems. This latter aspect forms the theme of a chapter on structure determination by physical methods, which attempts to cover the principal methods used for studying macromolecular structure in solution and the solid state. Although the account is comprehensive, the length and depth of treatment of the various techniques seem wildly disproportionate to their relative importance. While I must admit to some prejudice in the matter, surely X-ray crystallography has told us more about structure at the molecular level than any other approach; was it really a conscious decision of the editors to devote less than 15 pages to this method out of a 165-page chapter, which includes, for example, 30 pages on action spectroscopy, an approach of much more limited applicability? The excessively condensed account of crystallography permits no effective discussion of fibre diffraction or of the ways in which crystal electron-density maps are interpreted, matters that are essential for an appreciation of the significance and limitations of crystallography. Electron microscopy is covered well (including image reconstruction methods), light microscopy not at all. Raman spectroscopy (of increasing interest nowadays) is dealt with but briefly, infra-red spectroscopy merely *en passant* in another chapter.

Similar flaws are evident elsewhere in the book. A discussion of mechanisms of energy transfer gives too little emphasis to their roles in biological systems. And while the chapter on radiation biophysics is practical and clear, that on isotope methods could well have given more examples, such as the elucidation of enzyme mechanisms and the mobility of membrane proteins. Likewise, although much of the chapter on energetic and statistical relations is clear and solid, parts of the topic are covered in disproportionate depth. The chapter on enzymes as biological catalysts is, in contrast, far too scanty a coverage of a field to which biophysics has made contributions of enormous significance. No example is treated in detail, there is no consideration of the ways in which crystallography and spectroscopy have provided complementary views of structure-function relationships and there is no discussion in

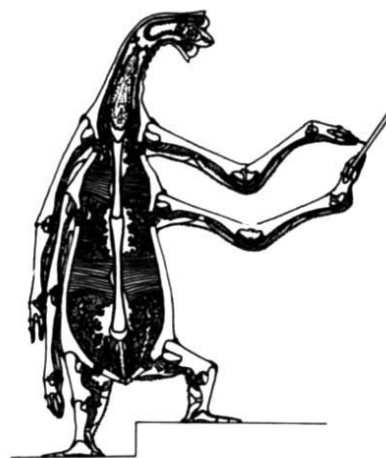
structural terms as to how biological control is effected. The best-understood example of this at the present time, haemoglobin, is inexplicably relegated to a brief mention towards the end of the long and generally good chapter on membranes.

Later chapters are concerned with photobiophysics, biomechanics (a pity here that the authors missed the opportunity to relate mechanical properties to the underlying macromolecular construction, for example of bones) and neurobiophysics. The latter chapter, generally both helpful and comprehensive, illustrates the problems of multi-authorship by a brief description of bacteriorhodopsin on page 713 that is neither in the index nor related to the earlier coverage of the topic on page 132; more surprising is that the elucidation of the structure of this protein, a *tour de force* of scientific investigation, is not referred to in the section on image reconstruction electron microscopy.

The penultimate chapter, on cybernetics, uses unfamiliar jargon to describe information theory, is highly specialized and becomes increasingly speculative. Finally, in a chapter on evolution, the editors have given three authors an opportunity to speculate from different standpoints on the origins of life and the ways in which present-day life forms could have evolved. While many points in this chapter are undoubtedly stimulating, its length of one hundred pages gives it a weight and appearance of validity out of all proportion to its status.

Biophysics was originally written in German and the version reviewed is an English translation of the second edition. It was unwise of the editors to give each author the option of using a professional translator or of rendering his own text into English. Sentences such as "not until

Spring books



Next week's *Nature* includes the Spring books review supplement. Over thirty books will be reviewed, among them A.K. Dewdney's *The Planiverse* from which the illustration above (of an Ardean, a two-dimensional being) is taken.

relatively late learnt one to build consistent models" or "this was, however, formulated not before 1855 by Virchow" are not uncommon. Furthermore, it is nothing less than a disgrace that there are so many errors in the rendering of German words into English — cybernetics as cybernetic, trypsin as trypsine, function as funktion, genome as genom among many other examples.

So, have the authors of *Biophysics* given us the definitive statement that we have been awaiting? Sadly, the book is too flawed to be acceptable in itself. If we try to forget the often appalling English, we cannot forgive the almost perverse way in which the authors have been allowed to exaggerate or ignore aspects of the subject in such an arbitrary fashion.

In this review I have drawn attention to the shortcomings of the book; unhappily, *Biophysics* falls a long way short of my hopes. It is, nevertheless, a remarkable achievement. The book contains a wealth of information, much of which is presented outstandingly well. It does indeed provide a definition of the scope of biophysics such as we have not had before — and the authors' work will be of the most inestimable benefit to their successors who seek to write a balanced and comprehensive text either for undergraduates or for research workers in biophysics. □

Anthony C. T. North is Professor in the Astbury Department of Biophysics, University of Leeds.



AMERICAN NUCLEAR SOCIETY makes available the ANS Edition . . .

Nuclear Power and Its Environmental Effects

by Samuel Glasstone and Walter N. Jordan; 408 pages; hardbound \$26.95; softbound \$18.95.

Mail to:

**American Nuclear Society
555 N. Kensington Avenue
La Grange Park, IL 60525 USA**

Name _____

Address _____

City _____

State/Province _____

Country _____ Zip _____

Please take my order for _____ books.

☐ Payment in full enclosed.

☐ Bill me and I accept postage and handling charges.

Seeing heresy in its true colours

J.D. Mollon

A Cybernetic Approach to Colour Perception.

By N.C. Paritsis and D.J. Stewart.
Gordon & Breach: 1983. Pp.167. \$29.50.

COLOUR vision attracts more than its fair share of books by amateurs, by eccentrics and by outright cranks. All of us — except the rare individuals who are completely colour blind — daily experience a vivid and elaborate range of hue sensations; and perhaps it is this subjective immediacy of colour, and the evident structure of our internal palette, that leads so many gallant amateurs to advance theories of colour vision.

But those who review books on colour vision have good reason not to dismiss too lightly the writings of newcomers to the field. The reason is George Palmer. Definitely an amateur and definitely eccentric, Palmer in 1777 published his *Theory of Colours and Vision* in which he correctly suggested that all our sensations of hue depended on just three kinds of retinal receptor and that the sensation of white arose when all three types of receptor particle were thrown into motion. In later monographs this prescient but mysterious man suggested that colour blindness resulted from the impairment of one or more of the receptor types and that complementary colour after-images arose from differential fatigue of the three types. All this, some 25 years before Thomas Young gave his celebrated Bakerian lecture and commended the trichromatic theory to the Royal Society.

Nevertheless, despite the sobering example of George Palmer, I must be direct and declare that the present book is amateur, is distinctly eccentric and can be safely passed over by specialists in the field.

The early chapters of *A Cybernetic Approach to Colour Perception* offer a pedestrian history of colour science and an introduction to model neurones. The authors then proceed to the first part of their theoretical work: the spectral sensitivities of Lateral Geniculate cells are quantitatively reconstructed from transformations of the spectral sensitivity curves of the photoreceptors. By allowing themselves some elaborate transformations, the authors obtain some very good fits to the classical electrophysiological records of De Valois. But the precision they attempt is quite inappropriate in view of the photoreceptor data

with which they begin. To obtain spectral sensitivity curves for the receptors, they average together the densitometric measurements that Rushton disowned in 1964 and the early microspectrophotometric data, for macaques and man, published in 1964. They ignore species differences, and they give no details of the corrections (if any) that they have made for absorption by lens and macular pigment and for variation in the retinal region that is stimulated.

The second part of Paritsis and Stewart's theory is qualitative. They propose the existence of single cortical cells that are specific for the seven hues — red, orange, yellow, green, blue, indigo, violet — that Newton's assistant distinguished in the spectrum. Mutually inhibitory connections exist between the neurones specific for different hues. To account for colour constancy and contrast phenomena, the authors postulate a small number of special cells, called "adaptors", that collect information about average luminance and average chromaticity across the visual field; the adaptors feed forward and adjust the sensitivities of more central units that are specific to the colour of stimuli in a local retinal region.

There is the germ of a good idea here. But the authors finally hang themselves when (p. 119) they use reciprocal inhibition between pairs of central neurones to explain why grey or white is seen when lights of complementary colour are mixed. It is a tenet of modern colour theory that the results of colour mixing can entirely be explained in terms of the quantum catches in the three classes of retinal cone. If two lights are of different spectral composition but produce the same set of quantum catches, then they will look alike to the observer. Complementary lights are simply lights that, when mixed in the correct proportion, yield the same set of quantum catches as does white light. At all subsequent stages the neural signals are the same as those produced by white light and, within the context of the authors' theory, none of the seven types of central hue-coding neurone will be active. It is always good that there should be those who question the basic dogmas of any scientific field; but here it is clear (see p. 120) that the authors do not realize the seriousness of their heresy and are either confused or ambivalent in their espousal of it.

If my judgement of this book is harsh, the authors may take comfort in the spectacular error of one eighteenth-century reviewer. Most journals ignored George Palmer's monograph of 1777, but one review did appear in the *Monthly Catalogue* for that year. It runs to a single line, and is reproduced below.

**Art. 49. *Theory of Colours and Vision*. By G. Palmer. 8vo.
1 s. Leacroft. 1777.
A visionary theory without any colour of truth or probability.**

J.D. Mollon is a Lecturer in Experimental Psychology at the University of Cambridge.

Strategies of proof

Graham Higman

The Classification of Finite Simple Groups. Vol. 1, Groups of Noncharacteristic 2 Type.

By Daniel Gorenstein.

Plenum: 1983. Pp.487. \$59.50, £45.80.

THE completion of the classification of finite simple groups is one of the most exciting events of twentieth-century mathematics. The problem has intrinsic interest: finite symmetry groups are apt to crop up almost anywhere, and the simple groups are the blocks with which they are built. Thus questions about finite groups arise naturally. Knowing what the simple groups are does not necessarily trivialize them, but it does improve dramatically our chance of being able to answer them.

The list of finite simple groups is quite complicated. It has been known since the nineteenth century that the simple Lie algebras comprise four infinite series and five exceptional algebras. Each simple Lie algebra gives rise to a series of finite simple groups, one for each prime power. But, much as complex Lie groups can have more than one real form, sometimes there is more than one such series; and for more remotely analogous reasons, there are still further series when the prime power is 2^n or 3^n . The members of this whole assemblage are called groups of Lie type. In addition, there are two further series, the cyclic groups of prime order and the alternating groups, and twenty-six odd men out, the so-called sporadic groups. The groups in the list are not only all formally different; they often have very different flavours. For instance, the groups of Lie type can be studied and are best studied by the powerful abstract and general methods of algebraic geometry, whereas the sporadic group have an awkward individuality as well as an intricate beauty, and never quite fulfil the promise they seem to give of something more general behind them.

Discovering and describing these simple groups is much less than half the battle. It has then to be proved that there are no others. Any such proof must somewhere reckon with each group in the list, and so can hardly be much less complicated than the list itself. Furthermore, the proof is put together from the contributions of a large number of mathematicians, working over several decades. In the early years, the later course of the argument could not be foreseen, so that papers written then are not always tailored to the needs of their successors. Key concepts, essential to the completion of the proof, often arose late in the day, and the significance of earlier work had to be revalued in the light of them. And, of course, there have been false trails, and seductive by-ways, and misunderstandings, and even mistakes. The proof is, indeed, there in the literature, but it can

scarcely be said that it satisfies the requirement of accessibility to checking which is the mathematician's equivalent of the repeatability of experiments.

This is the situation to which Daniel Gorenstein has addressed himself in three books, of which the volume under review is the second, and which together seek to expand the whole proof, as it now is, in an accessible form. The first, entitled *Finite Simple Groups*, published in April 1982, described the groups in the list, gave an overview of the proof of completeness and developed the key concepts. The two remaining volumes put in the detail. It must be emphasized that, in spite of the length of the books, this is not detail for the sake of illustration or to clear up subsidiary questions; it is detail without which the proof would not be a proof. The method is determinedly historical, the expressed intention being to expound the original proof, with the improvements up to early 1982. Every step on the way is given the appropriate reference, so that the work constitutes a historical record as well as a mathematical exposition.

The proof of the classification theorem falls naturally into two parts. It has been clear since the pioneering work of Richard Brauer, and even more since the odd order paper of Walter Feit and John Thompson,

that the key to the classification is the 2-local structure, the structure, that is, of the normalizers of subgroups of order a power of 2. From this point of view, the groups of Lie type fall into two classes, according to whether the prime power involved is odd or even; and the distinction can be extended to other simple groups. This volume deals with the odd case; the second will deal with the even.

It cannot be doubted that Daniel Gorenstein was the right man to do this job. His personal contribution to the detailed work puts him among the top half-dozen workers in the field; but for generalship, for his early grasp of the wider strategies of the proof, he was quite outstanding. This book deserves the highest praise. It is not the last word on the subject, however; further rationalization of the existing proof is already on its way, and someone, perhaps, will come up with something quite new. And, no doubt, group theorists and users of group theory will go on using the classification theorem without knowing how to prove it. When the final volume appears in 1985 or 1986, they will no longer have any excuse. □

Graham Higman is Waynflete Professor of Pure Mathematics at the Mathematical Institute, University of Oxford.

Halley watching

David W. Hughes

International Halley Watch Amateur Observers' Manual for Scientific Comet Studies.

By Stephen J. Edberg.

Enslow Publishers, Hillside, NJ / Sky Publishing, Cambridge, MA: 1983. Pp.190. Pbk \$9.95.

OBSERVING Halley's comet in the winter of 1985 and spring of 1986 is going to be akin to ornithology. No serious bird-watcher would expect to leave his house and see a Wilson's Phalarope sitting on his gate post. To observe such an exotic species requires some effort: journeys to possible habitats, consideration of season, care over choice of observing instrument and so on.

Halley's comet at this apparition is going to be similarly hard to find and distinctly unstartling. Do not expect to walk outside and be amazed by a brilliant comet with a tail stretching from horizon to zenith. The geometry is unfavourable. When the comet is at its most active the Earth will be 1.5 AU away on the other side of the Solar System. The coma magnitude will barely creep above the level visible with the naked eye for people in highly populated regions.

To see Halley's comet one must be well prepared and Edberg's *Manual* is an indispensable accessory. It has been written for the advanced amateur astronomer so it does not give guidance on basic observing

techniques or methods of data reduction, but concentrates on explaining how the observer can generate meaningful scientific data. Part 1 introduces the International Halley Watch — the organization which will co-ordinate the observations and will eventually catalogue the data. Also stressed are the importance of visual observations of the coma and tail. This is one of the only techniques that can link present-day observations to those of many of the previous apparitions of the comet. The importance of dark adaptation, atmospheric transparency and sky brightness are emphasized. Photography (concentrating on the choice of emulsions, development processes, exposure times and lens focal length), astrometry (the accurate recording of the comet's position), spectroscopy, photoelectric photometry and observations of the two associated meteor streams are also covered.

Part 2 consists of an ephemeris (which, among other things, lists right ascension, declination, magnitude, heliocentric and geocentric distance at daily intervals between June 1985 and May 1987), a set of 19 star charts showing the movement of the comet against the sky background, and five figures showing the elevation and azimuth of the comet for observers at latitudes 40°N, 30°N, 20°N, 20°S and 30°S.

No amateur or professional astronomer who has the vaguest interest in comets should be without this book. □

David W. Hughes is a Lecturer in Physics and Astronomy at the University of Sheffield.

Faces of hominid evolution

Peter Andrews

The Australopithecine Face.

By Yoel Rak.

Academic: 1983. Pp. 169. \$45, £33.

IT IS curious how different lines of evidence assume greater or lesser importance at different times in our search for evidence on human ancestry. Teeth used to be considered of the utmost importance, then the postcrania, but most recently it has been the face. In this sense Yoel Rak's book on the australopithecine face is timely. It will set new standards for the comparative description of fossil hominids, both because of the breadth of the study and because of the rigour of the analyses of facial characters.

The face is one of the two main components of the skull (the other being the vault). It is described here for the species of *Australopithecus*, the group of fossil hominids most nearly related to our own genus, *Homo*. The essence of the book is the description, partly comparative, of each species of australopithecine, which is followed by interpretation of the morphology on functional and — to a lesser extent — phylogenetic grounds. It is shown that, contrary to widespread belief, the face of *A. africanus* is extremely specialized and widely divergent from the hominoid ancestral pattern. Far from providing a good model for human ancestry, it is in many respects very different from the common pattern shared by *Homo* and the African apes, the chimpanzees and gorillas. Moreover, many of these characters are shared with the robust australopithecines, especially *A. robustus*, and the third species, *A. boisei* (formerly *Zinjanthropus*), is interpreted by Rak as even more derived.

The similarities in structure between the species of *Australopithecus* are interpreted functionally in a novel but plausible way. What is really striking about this book is the way the functional changes are shown to be the products of the different combinations of characters. For instance, the buttressing of the face, represented by several distinctive characters, is related to facial projection and tooth size. This trend begins in *A. africanus* and is carried to its functional extreme in *A. robustus*, while in *A. boisei* there is another trend leading to advancement of the peripheral parts of the face associated with retractions of the lower face and reduction of central buttressing. Whether these trends are seen as sequential, which is Rak's view, or whether they could occur in parallel, with *boisei* and *robustus* as independent offshoots of an *africanus*-like morphology, is not yet clear.

So far I have only mentioned three species of australopithecine, but Rak also

describes the more fragmentary remains of the fourth species, *A. afarensis*. No complete face or skull of this species is known. A considerable proportion of the face can be reconstructed from several less complete specimens, however, and the interesting thing here is that none of them show any sign of the functional complexities described for the other three species of *Australopithecus*. By comparison with other hominoids, especially the African apes and *Homo*, Rak interprets the *A. afarensis* condition as primitive and the other species of *Australopithecus* as derived. I think he demonstrates this well. What he does not show, or even question, is why he puts *afarensis* into the genus *Australopithecus* at all. If it lacks the derived characters that are present in the other species, why group it with them? Based on the plesiomorphous characters of its face, but with clear adaptations to bipedality, it would seem more likely that *afarensis* is the sister group to both *Australopithecus* and *Homo*, the position accorded to it by its describers, in which case of course it could not be referred to either genus without making that genus paraphyletic.

There is no doubt, however, that this book will be central to discussions on human evolution for many years to come. It contains a wealth of information that is well described, well analysed and well interpreted. It will be an essential acquisition for all libraries and all individuals interested in human evolution. □

Peter Andrews is in the Department of Palaeontology at the British Museum (Natural History), London.

One Piaget

M. Scaife

Piaget: Critique and Reassessment.

By David Cohen.

Croom Helm/St Martin's Press: 1983.

Pp. 163. Hbk £13.95, \$25; pbk £6.95.

A WORK on Piaget which concludes that "this really ought to be the last book to assess him as a contemporary psychologist" is unlikely to be much concerned with the problems of constructing adequate developmental theories. David Cohen's presentation of Piagetian theory supports this surmise. Partly because Piaget invoked work from several disciplines, there is some confusion as to what label he should be given: was he "really" a psychologist for instance? Piaget's own answer was that he was a genetic epistemologist and that this term defined his work. Cohen does not seem to have grasped the significance of this point and thus fails to comprehend the distinction between the "epistemic" and the "psychological" subject as an object of

investigation. This leads to the bizarre criticism that, in contrast to Freud's vivid descriptions of individual patients, Piaget "never leaves one with a feeling that he has given any portrait of [an individual] child" (p.82). This may be bad journalism on Piaget's part but it is scarcely a fault from the viewpoint of genetic epistemology.

Another crucial relationship in Piagetian theory that the author has trouble discussing is that between the biological and the psychological "realm" of explanation, especially between phylogenetic and ontogenetic processes, a link that Piaget captured in the phrase "Intelligence is an adaptation". Here we are told that "one of [Piaget's] key suggestions was that the development of the child's mind recapitulated the development of the species as a whole. Modern four year olds might, therefore, think like adult Neanderthals" (p.78). But this is a misrepresentation of Piaget's view and is exactly the kind of simplistic argument that he specifically cautioned against. What hope, then, is there for understanding Piaget's formulations on the relationship between structure and function or on the concepts of equilibration, homeorhesis, the phenocopy and so on? Cohen solves the problem by not mentioning them, perhaps a blessing since the book is already replete with inexact usages of terms. The needless misuse of "egocentric", for example, in common with so many other authors, makes one see clearly why Piaget eventually felt forced to stop using the word.

As a useful critique, then, this book is of little value and it is unsurprising that the omissions are serious. There is no mention of critics such as Pierre Mounoud or Anthony Wilden. Even Henri Wallon, who spent years in debate with Piaget, gets but a single line.

Instead of serious theoretical argument, Cohen offers a mixture of largely familiar experimental data presented in an astonishingly patronizing manner: Piaget "need not be blamed for not being the perfect experimenter of today" says Cohen (p.97) since his work occurred before we knew that subjects may be confused or act irrationally or to please the experimenter. Nor was there anyone to point out the error of his ways for Piaget, we are told, was "le patron", the undisputed boss, a Freud without the fractious disciples egged on by their egos" (p.81). Cohen's naive and cosy vision of Genevan life is supported by long quotations of honeyed phrases from special Piagetian issues of journals and obituary notices. All of this is intended to lead the reader to conclude that Piaget's genetic epistemology is nothing more than a second-rate developmental psychology. Yet the most dispiriting aspect of this volume is that such superficiality will find favour with so many psychologists. □

M. Scaife is a Lecturer in Psychology in the Cognitive Studies Programme, University of Sussex.

Dishes to dispensers

Sixteen new products are described this week, ranging from a dish for the reception of data from weather satellites to an information package on liquid chromatography techniques.

●A low-cost system to receive weather images from orbiting and geostationary satellites has been launched by Feedback Instruments of Crowborough, England. The WSR513 comes as a complete package and provides real-time images of weather patterns in most parts of the globe. The cost has been kept down by careful circuit design. Orbiting satellites provide real-time images of the part of the globe over which they are passing provided they are also within range of the receiver's antenna. Geostationary satellites give sequential images of fixed sectors of the Earth's surface. The basic system will receive VHF signals from the APT (Automatic Picture Transmission) systems of the US National Oceanic and Atmospheric Administration's Tiros and Meteor series of orbiting satellites. Addition of a parabolic dish aerial and extra modules allows transmissions from geostationary satellites such as Meteosat, which covers a large area of Europe, to be received. The transmissions from Meteosat are also augmented from time to time by information from GOES Central, the geostationary satellite which relays weather data for North and South America and much of the Western Atlantic.

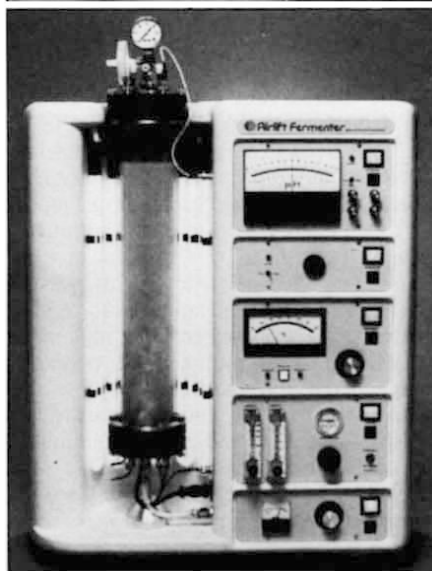
Circle No. 135 on Reader Service Card.

●The Beckman Instruments pH131 is a portable laboratory bench model pH meter which makes use of a custom-designed integrated chip. The meter automates most measurement steps to provide nearly unattended measurement and simple operation for routine tasks such as quality control and environmental monitoring. It has display, error and out-of-temperature-range flags; always displays temperature; and does temperature computation with or without an automatic temperature compensation probe.

Circle No. 136 on Reader Service Card.

●The Isolab Glyc-Affin kit for measuring glycosylated haemoglobins by affinity chromatography is a possible replacement for the standard ion-exchange methods, and can also measure other glycosylated proteins. The columns, from Advanced Laboratory Techniques of Tunbridge Wells, contain a boronate agarose medium that has an affinity for all glycosylated proteins. It can be used to measure glycosylated haemoglobins with assay results comparable to those obtained by the Fast Hb test. In addition, the Glyc-Affin System will quantitate glycosylated serum proteins.

Circle No. 137 on Reader Service Card.



●Dionex 2300i series instruments combine Dionex ion chromatographs with AutoIon 300 software and high performance Hewlett Packard computer hardware to provide a completely interactive, computer-automated ion analysis system.

Circle No. 138 on Reader Service Card.

●The BRL Airlift fermenter is a bench-top fermentation system able to create precise reproducible growth conditions for a wide variety of microorganisms, including aerobic, anaerobic, thermophilic and photosynthetic organisms. The self-contained fermenter is complete with fully-controllable lights, air/gas mixing and delivery capability, built-in air compressor. The instrument panel contains controls for pH, light intensity, temperature (to 121°C), and custom gas mixing, as well as monitoring foam and fluid levels.

Circle No. 139 on Reader Service Card.

●Measurements of heats of dissolution, heats of hydration, heats of mixing and heats of reaction can be performed in a new calorimeter from Astra Scientific of California. Other measurements such as specific heat of solids or liquids and phase transitions are also possible. Application areas include the study of wetting characteristics of materials; determining the solubility of compounds; and evaluating reaction kinetics and synthesis.

Circle No. 140 on Reader Service Card.

●M13mp18 and M13mp19 are the newest addition to the Pharmacia P-L line of products for M13 cloning and sequencing. These wild-type vectors can be propagated on *E. coli* JM105, which lacks restriction and modification systems. The use of M13mp18 and M13mp19 with this host eliminates the problem of degradation of large unmodified DNA fragments that can occur in the JM103 host used with previous M13 vectors. M13mp18 and M13mp19 also have unique *Sph* I and *Kpn* I sites in the multiple cloning site.

Circle No. 141 on Reader Service Card.

●Knauer's Model 50 microprocessor programmer is designed to complement the company's dual reciprocating pump system. A back-up battery stores up to ten independent programs for simple isocratic operation. The programmer can also be used for other laboratory instrumentation and for general automation processes. Auto samplers, injection valves, column switching valves, solenoid valves, and many other types of auxiliary instruments can be easily connected and controlled without extensive programming experience. An advanced version (Model 50A) adds two printed circuit cards to expand to a ternary system, the programmer acting both as a lab timer and a gradient former. Solvent composition can be programmed for any time span, giving a linear change of composition.

Circle No. 142 on Reader Service Card.

●The Brinkman digital burette is a bottle-top burette with an LCD digital display to avoid the risk of meniscus reading errors. The display, which reads to 0.01 ml, indicates titrated volumes up to 99.9 ml, then automatically re-zeros and begins counting again.

Circle No. 143 on Reader Service Card.

●Acro LC3 from Gelman Sciences is a disposable 3mm filtration device for filtration of aqueous and solvent-based HPLC samples. Designed for filtration of samples in small volumes, the Acro LC3 features a female luer inlet for attachment to standard syringes. The filter comes in two models both with 0.45µm pore size. The LC3S is designed for filtration of solvent-based samples and contains a 3mm PTFE membrane for solvent resistance. The LC3A, for aqueous solutions, contains Versapor, a non-particulate-releasing, hydrophilic membrane.

Circle No. 144 on Reader Service Card.

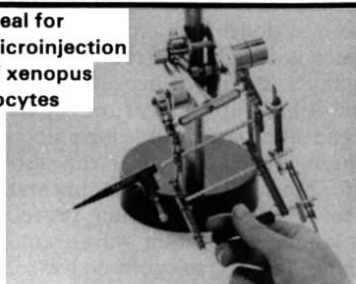
●The BioVortexer mixes liquid samples in test tubes without the hand manipulation required with a conventional vortex mixer. Centrifugal pellets, precipitates and other solids are readily dispersed by a combination of vigorous mechanical and vortex action. Cross-contamination is avoided by using throw-away, polypropylene stirring tubes. The BioVortexer is battery driven.

Circle No. 145 on Reader Service Card.

ADVERTISEMENTS

**FAIRCREST
GLOVE BOXES**
Barretts of Croydon
1 Mayo Road
Croydon,
CR0 2QP U.K.
Tel: 01-684 9917

Ideal for
microinjection
of xenopus
oocytes

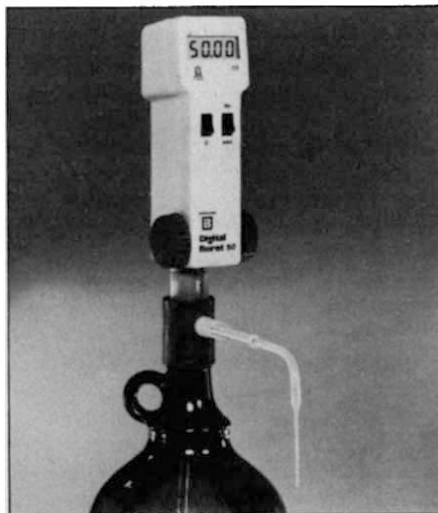


SINGER MICROMANIPULATORS ARE FAMOUS BECAUSE

Singer Micromanipulators enable anyone to make exact movements without tremor or lost motion, in three dimensions, using a single control handle under low or high power magnification.

A varied range available. All simple to operate.

SINGER INSTRUMENT CO. LTD.,
Treborough Lodge, Roadwater, Somerset.
Tel: Washford (0984) 40226/325



The Brinkman digital burette and the Pharmacia information package.

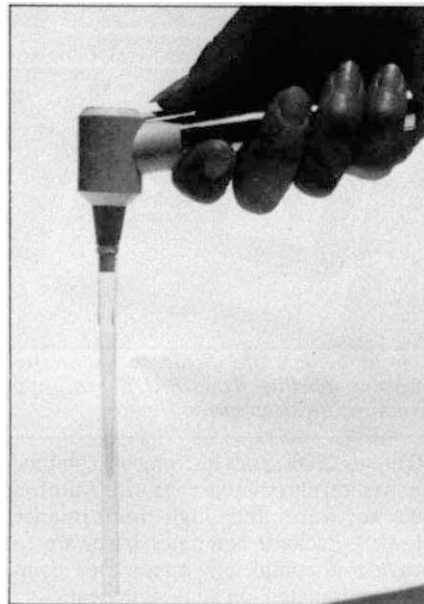


●A literature package on alternative standard techniques for liquid chromatography has been compiled by Pharmacia Fine Chemicals. A standard technique is defined as a liquid chromatography method with a wide applicability. Established standard techniques are gel filtration and ion-exchange chromatography. The package contains four data sheets with methodology for covalent chromatography, chromatofocusing, hydrophobic interaction and metal chelate affinity chromatography. The file is available free of charge from Pharmacia.

Circle No. 146 on Reader Service Card.

●Perlag capillary aggregate from Silvaperl is a novel plant growth medium produced from volcanic lava by a thermal process which gives each particle a highly porous internal structure that can hold up to half its own weight in water. The interconnected capillaries and cellular pores in Perlag make it free-draining, but at the same time give it built-in controlled moisture retention and release, making it an excellent benching medium.

Circle No. 147 on Reader Service Card.



For mixing in test tubes, the Bio Vortexer.

●What is claimed to be the first commercially available instrument for measurement of forces as low as 1 µg with good frequency response has been introduced by Cambridge Technology, Inc. of Cambridge, Massachusetts. The Model 406 force transducer system is based on a variable displacement capacitor whose plates are formed from an etched, cantilevered, quartz beam, with an integral reference element to simplify operation. The reference element provides good temperature stabilization, and simultaneously rejects most ambient vibration and electrical interference. The unit can measure up to 50 mg full-scale and its 100 Hz, 3 db down frequency response allows a critically damped step response of 5 ms (1 to 99%). The thin layer of air between the capacitor plates provides mechanical damping.

Circle No. 148 on Reader Service Card.

●The latest IQ 190 sample processor from Cavo Scientific Instruments is fitted with a programmable digital dilutor for applications in the clinical laboratory. The processor can be used for pipetting and diluting, titrating, serial diluting, dispensing, multiple dispensing, aspirating and aliquotting samples and/or reagents. The unit can aspirate a sample and up to 8 reagents for complicated procedures. An optional wash function eliminates contamination between samples.

Circle No. 149 on Reader Service Card.

●New from Biomedical Technologies is a sensitive cyclic AMP radioimmunoassay kit. The kit includes a highly specific pre-conjugated cyclic AMP antibody and an iodine-125 labelled tracer. The sensitivity of the assay is 0.10 picomoles per tube, or 0.005 picomoles per tube for acetylated samples. The protocol is simple, includes only one incubation and generates a visible precipitate.

Circle No. 150 on Reader Service Card.

These notes are based on information provided by the manufacturers. For further details circle the appropriate numbers on the Reader Service Card bound inside the journal.